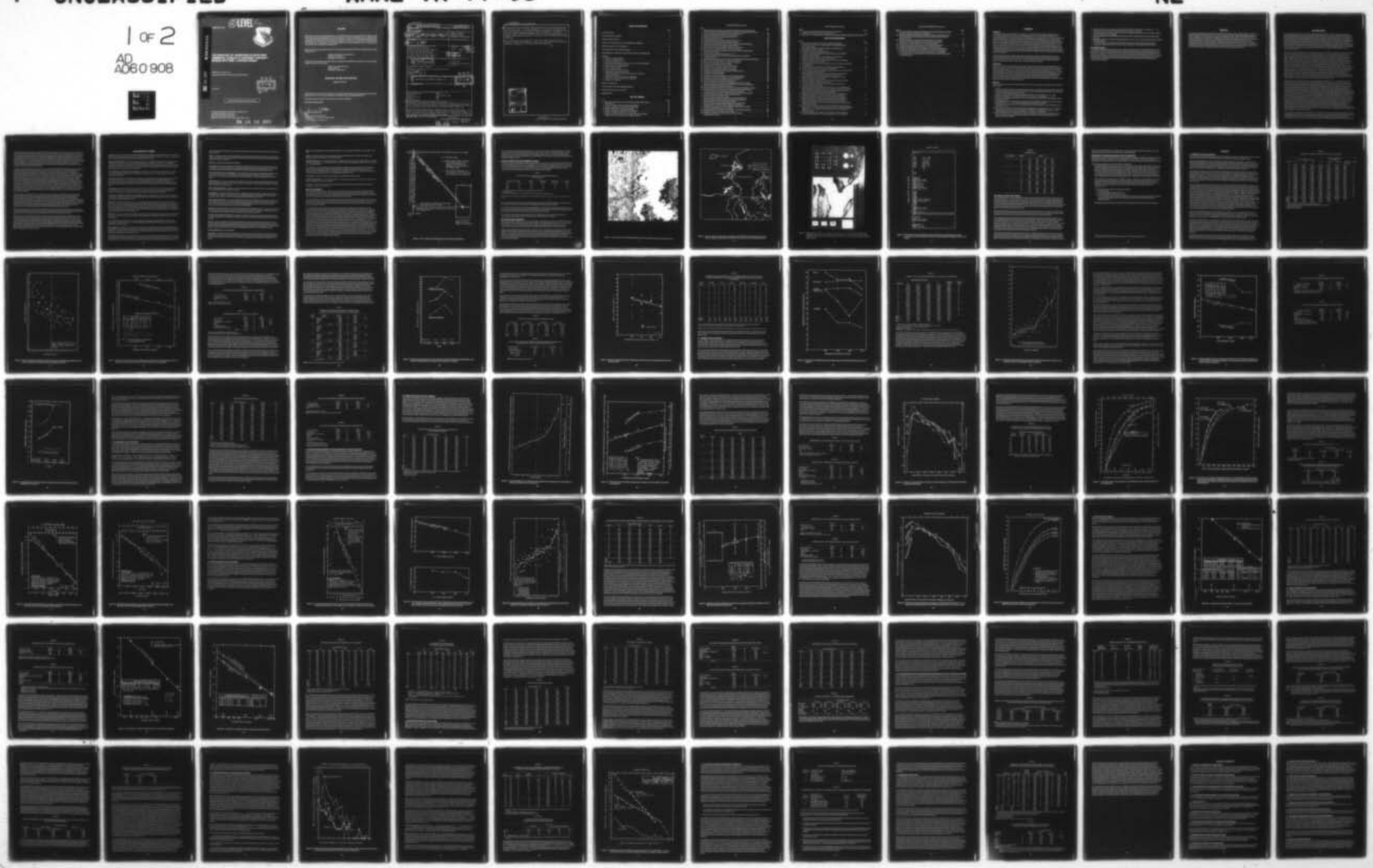
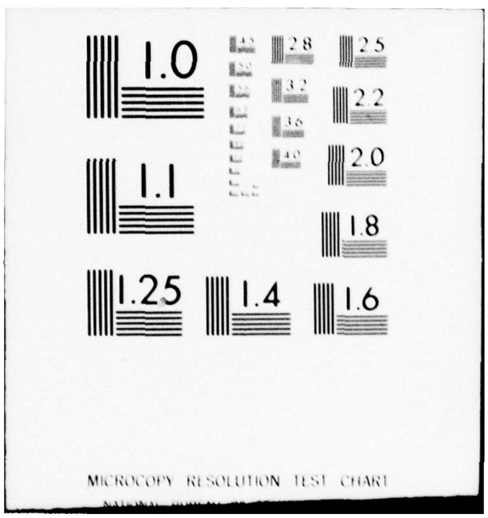
AD-A060 908 AEROSPACE MEDICAL RESEARCH LAB WRIGHT-PATTERSON AFB OHIO F/G 5/9
THE BEHAVIOR OF OBSERVERS IN DETECTING UNBRIEFED TARGETS AT DIF--ETC(U)
JUN 78 H C SELF
UNCLASSIFIED AMRL-TR-77-95 NL

1 OF 2
AD
A060908




1.0
2.8
2.5
3.2
2.2
3.6
4.0
2.0
1.1
1.8
1.25
1.4
1.6
MICROCOPY RESOLUTION TEST CHART

AMRL-TR-77-95

THE BEHAVIOR OF OBSERVERS IN DETECTING UNBRIEFED TARGETS AT DIFFERENT AIRCRAFT SPEEDS WITH SIDE—LOOKING RADAR

HERSCHEL C. SELF, Ph.D.
AEROSPACE MEDICAL RESEARCH LABORATORY

JUNE 1978

AEROSPACE MEDICAL RESEARCH LABORATORY
AEROSPACE MEDICAL DIVISION
AIR FORCE SYSTEMS COMMAND
WRIGHT-PATTERSON AIR FORCE BASE, OHIO 45433

NOTICES

Please do not request copies of this report from Aerospace Medical Research Laboratory. Additional copies may be purchased from:

National Technical Information Service
5285 Port Royal Road
Springfield, Virginia 22161

Federal Government agencies and their contractors registered with Defense Documentation Center should direct requests for copies of this report to:

Defense Documentation Center
Cameron Station
Alexandria, Virginia 22314

TECHNICAL REVIEW AND APPROVAL

AMRL-TR-77-95

This report has been reviewed by the Information Office (OI) and is releasable to the National Technical Information Service (NTIS). At NTIS, it will be available to the general public, including foreign nations.

This technical report has been reviewed and is approved for publication.

FOR THE COMMANDER

CHARLES BATES, JR.
Chief
Human Engineering Division
Aerospace Medical Research Laboratory

AIR FORCE/56780/30 August 1978 — 100

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER (14) AMRL-TR-77-95	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle) (6) THE BEHAVIOR OF OBSERVERS IN DETECTING UNBRIEFED TARGETS AT DIFFERENT AIRCRAFT SPEEDS WITH SIDE-LOOKING RADAR.	5. TYPE OF REPORT & PERIOD COVERED TECHNICAL REPORT 11/10/64 - 11/10/77
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) (10) Herschel C. Self, Ph.D.	8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Human Engineering Division Aerospace Medical Research Laboratory Wright-Patterson AFB, Ohio 45433	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 62202F 7184 04 (17) 04
11. CONTROLLING OFFICE NAME AND ADDRESS Aerospace Medical Research Laboratory Aerospace Medical Division, AFSC Wright-Patterson AFB, Ohio 45433	12. REPORT DATE (11) June 1978
	13. NUMBER OF PAGES 103
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) (16) 7184, 665A	15. SECURITY CLASS. (of this report) Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

(12) 103 p.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

(9) Technical rept. 10 Nov 64-10 Nov 77,

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Human engineering	Radar
Detection	Side-looking radar
Target detection	SLR
Targets	SLAR
False positives	

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The numerous false positives found in SLR studies with unbriefed targets pose a severe problem for operational systems. The present study examines this problem and derives mathematical equations for describing observer behavior.

Twenty USAF Radar Observers were trained and tested on side-looking radar. They searched for unbriefed airfields, industrial sites, railroad yards and tank farms at simulated aircraft speeds of 700-2110 knots. A data camera photographed every object on the display designated as a target.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Unclassified
SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

Tripling aircraft speed reduced target detections by only 17% while reducing reaction time by 56%. The high percentage of false positives was found to be due to the similarity of the radar signatures of targets and non-target objects. The false positive problem was shown to not be solvable by: (1) Better selection and/or training of observers, (2) Use of the expressed confidence in response correctness of observers, or (3) Use of teams of independently-working observers.

The relationships between measures of performance were examined in detail. Selection of superior observers was found to be complicated by the conflicting behavioral requirements of different performance measures.

Unclassified
SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

TABLE OF CONTENTS

LIST OF TABLES

LIST OF TABLES (Continued)

LIST OF TABLES (Continued)

LIST OF ILLUSTRATIONS

LIST OF ILLUSTRATIONS (Continued)

SUMMARY

PROBLEM

This report describes an experimental investigation of various aspects of the target-finding behavior of side-looking radar observers searching for unbriefed targets over a wide range of aircraft speeds. It was conducted to answer several questions about target-finding behavior. Among them are: (1) How do measures of observer performance vary with aircraft speed? (2) To what extent are different measures of performance related? (3) Can simple mathematical equations be derived that accurately describe some aspects of observer behavior? (4) How large are the differences in performance measures of side-looking radar observers? (5) Is it possible to select observers who are superior on most measures of performance or does superiority on one or more measures go along with inferiority on others? (6) Can independently-working teams of radar observers do better than lone observers? (7) Are any cues to solving the problem of excessive numbers of false positives (nontargets mistaken for targets) of side-looking radar observers apparent from examination of the images on the display of targets and false positives? (8) Does a measure of response similarity yield insights into selection of superior observers?

A series of previous reports on side-looking radar by the author and his co-workers had raised the above questions, and had shown that the utility of side-looking radar for finding unbriefed targets was seriously limited by two aspects of observer behavior: (1) Observers did not detect a large percentage of unbriefed targets. (2) Observers mistake a large number of nontarget objects for targets.

APPROACH

Twenty U.S. Air Force radar navigators were given 10 hours of intensive training in finding targets with side-looking radar. Using a Latin square experimental design, observers performed the detection task at four different aircraft speeds: 700, 1170, 1640, and 2110 knots. During testing they observed a 14" x 14" display screen showing a 10 nautical mile square of territory from a film strip covering 540 nautical miles of territory. Aircraft speed was simulated by smooth continuous motion along the film strip. The unbriefed targets that they tried to find, 78 in number, were airfields, dams, industrial sites, railroad yards and tank farms. All objects mistaken for targets were marked on a second copy of the filmstrip for later study by the author. Measures of performance included numbers of targets detected, numbers of false positives, reaction time, and derivations and combinations of these measures.

RESULTS

Experimental results are described and discussed in detail in the remainder of this report. The main findings are as follows:

1. Differences between observers on all performance measures exceed differences at different aircraft speeds.
2. Tripling aircraft speed reduced detections by 16% and decreased target detection time by 56%: Observers work faster and detect almost as many targets.
3. Number of targets detected decreases linearly with increase in aircraft speed, V: $N = A - BV$.
4. Number of false positives also decreases linearly with increase in aircraft speed: $n = C - DV$.
5. Detectivity of unbriefed targets was low: For no type of target at any speed did the percentage of targets detected exceed 30%.
6. Ground distance traveled, S, between the appearance and the detection of targets was linearly related to aircraft speed, V: $S = A + BV$. Tripling aircraft speed increased S by only 30%.
7. The percentage of targets detected, P, was exponentially related to distance, X, down the display when detected: $P = e^{A+BX}$.
8. The average time to detect targets, $\bar{t}$ decreases as the logarithm of aircraft speed, V: $\bar{t} = B - A \operatorname{Log}(V)$.
9. A large portion of objects mistaken for targets have "signatures" (images) more like those of "good" targets than does the average real target. Observer training is not the problem.
10. Teams of independently-working observers using various decision rules on what would be counted as targets were able to only slightly reduce false positives, and that at a cost of a drastic reduction in percentage of targets detected.

11. Numbers of detections and numbers of false positives for observers are positively correlated ($r = +.67$): Those observers who are "good" on one measure tend to be poor on the other.
12. Those who detect more targets tended to have a lower percentage of false positives even though the actual number of false positives was higher.
13. Speed of response was not significantly related to percentage of targets detected nor percentage of false positives.
14. Expressed confidence in correctness of response had little, if any, value in discriminating between targets and false positives.
15. An index of response similarity showed that observers who detect many targets tend to find many "unpopular" targets and nontargets, and those who mistake many nontargets tend to do likewise.

RECOMMENDATIONS

(1) In missions seeking targets of opportunity very high aircraft speeds can be used with little loss in observer performance. (2) On some measures of performance some radar observers are much better than others, but those who are superior on most measures are rare. Observers for specific missions should be selected to meet specific mission objectives. (3) The twin problems of low detectivity of unbriefed targets and excessive numbers of false positives with side-looking radar do not appear to be solvable by either more training or by teams of independently-working observers. Solutions must be looked for in operator aids, auxiliary equipment or improved radar equipment.

PREFACE

This report was prepared in the Human Engineering Division of the Aerospace Medical Research Laboratory, Wright-Patterson Air Force Base, Ohio. The work was performed jointly under Program 665A, Precision Strike, and Project 7184, "Man-Machine Integration Technology," Task 718404, "Visual Processes in the Perception of Displayed Information." Special thanks are due to the Strategic and Tactical Air Commands for supplying officers to serve as test subjects. Thanks are due to Mr. Don F. McKechnie for training the experimental subjects to recognize targets on side-looking radar displays and to Ms. Barbara Van Ausdall Staples for assistance in testing the observers. The author thanks the Westinghouse Electric Corporation for supplying the side-looking radar pictures utilized to illustrate this document. Thanks are due to Mrs. Betty Reid, Mrs. Kathy Hauser, and Miss Patricia Allen for help in preparing the manuscript.

INTRODUCTION

Side-looking radar (SLR), sometimes called side-looking airborne radar (SLAR), is capable of displaying images of ground targets that are noteworthy because of their resolution and contrast. Images are frequently of such high quality that a trained SLR observer quickly recognizes many of them. However, well-imaged targets are sometimes undetected because the observer does not look directly at them: He is searching some other part of the display. In an unbriefed target situation, missions may not be successful even though the SLR provides good quality target images. When searching for unbriefed targets, sometimes called "targets of opportunity", past research has shown that observers provided with high quality target images do not quickly find most targets and mistake many nontarget objects for targets. The present paper examines various aspects of observer behavior related to target detection and quantifies some aspects of SLR performance with mathematical equations.

Most earlier research on target detection and recognition has done little more than point out some of the many variables that must be taken into account if prediction equations are to be formulated. Some of the earliest laboratory studies in aerial reconnaissance (Boynton and Bush, 1955 and 1957; Boynton, Elworth, and Palmer, 1958) used stimulus material that did not resemble terrain or real objects to examine the effects of image variables such as target size, contrast, brightness, and complexity of the target background. They found, as one would expect, that recognition performance improved with target size, contrast, exposure time and observer experience and decreased with increased number of confusional objects. Probability of correct response increased linearly with exposure time, decreased linearly with the logarithm of the number of displayed objects, and decreased linearly with subject-figure distance. On complex displays of randomly-drawn figures, a study done at Wright-Patterson Air Force Base (Baker, Morris and Steedman, 1960) obtained a significant positive correlation between detection time and the percentage of observers who misidentified a target. Both search time and errors increased with: (1) increase in the number of irrelevant forms on the display, and (2) increase in the difference between the resolution of the reference form and that of the displayed target. In another study, now a classic (Steedman and Baker, 1960), it was found that search time and errors with random form targets in a matrix of forms was invariant until the visual angle subtense of the target fell below 12 minutes of arc; below 12 minutes performance deteriorated.

A 1960 study on operator performance in strike reconnaissance (Williams et al.) varied the resolution of photographs of various image scales and allowed unlimited time to find objects. In part of the study time was limited in finding airfields. The authors concluded that prediction equations were possible. Conklin (1962) exhaustively examined the importance of target-background parameters such as shape complexity, pattern complexity and background complexity. Nygaard et al. (1964) did research on the influence of stimulus complexity and achieved some success in relating objective measures of target and background complexity to operator performance with SLR, infrared and aerial photographic images. Rhodes (1964) related judged image complexity to the time taken by observers to find targets in aerial photographs. He found that eight orthogonal factors accounted for 86% of observer variability in performance. A year earlier Roetling et al. (1963) had related the amount of intelligence information extractable from photographs by photo interpreters to contrast, grain, resolution and passband. However, only 13% of the variance in accuracy was accounted for by the four factors. These studies suggest that the information content in an image is proportional to complexity, which includes contrast, resolution, number of objects, shape of target versus other objects, etc. It has been shown that observer performance can vary widely when size, resolution and contrast of the target appear to be entirely adequate: The complexity of the image and the target-background relationships cause large performance variation. Adequate prediction of performance is possible only if background characteristics and target-background interaction are taken into account, along with more conventional physical characteristics, such as image size, resolution, contrast, etc.

In a pilot study on the effects of image motion rate on observer performance with SLR (Self and Rhodes, 1964), college students served as observers. Although image motion rates simulated aircraft speeds ranging from 600 to 2,000 knots, statistical tests showed that observer performance did not degrade with an increase in simulated aircraft speed. In a second study by the same authors (Rhodes and Self, 1964), the effects of direction

of image motion across the display at 635 knots and at 1780 knots were measured using radar operators from the Strategic and Tactical Air Commands as observers. At the slower simulated aircraft speed, significantly more targets were detected and the time between the appearance of targets on the display and their detection was significantly shorter. However, simulated aircraft speed had no significant effect upon the number of nontarget objects that were mistaken for targets. In a study concerned with target briefing and aircraft speed (McKechnie, 1967), a statistically significant 8% loss in number of targets detected occurred in going from 600 knots to 3000 knots, i.e., with a 5-fold increase in speed. However, briefing and speed were confounded. Elworth (1964) found that the number of missile sites acquired from photographic film strips was inversely related to the log of simulated aircraft speed.

In previous studies on targets of opportunity by the author (Self and Rhodes, 1964; Rhodes and Self, 1964; Van Ausdall and Self, 1964), the number of false positives was large. A high percentage of false positives in operational systems would impose severe restraints on their utility. For this reason it is worthwhile to examine in detail observer performance on false positives and the images of the objects mistaken for targets. This would aid in deciding whether or not more training or different training would be of value in solving the false positive problem, or whether additional sensors or other observer aids would be necessary. Possibly observers could be selected who would not respond to so many false positives, and/or procedures could be worked out to minimize the number of false positives. If SLR imagery contains many nontarget images that more closely resemble the images of well-resolved real targets than do the images of any real targets, then a second sensor, or some other operator aid, may be necessary to supplement the SLR sensor. Because of the magnitude of the false positive problem, the present study will examine and analyze the false positive data in considerable detail. In addition to examining the value of using selected observers to minimize the number of false positives, use of teams of independent observers having different decision rules will be examined for minimizing the problem.

Numerous studies by the author and his coworkers, as well as by other researchers, that have examined observers' ability to find and recognize targets on nonuniform backgrounds have found large differences between observers. Frequently the differences between observers have increased, not decreased, with increased training and experience. In addition, from one test session to the next many individuals vary greatly in performance. The large individual differences indicate that the selection of individuals for observer training schools and for duty as observers on missions is very important. The author (Self, 1972) had an earlier paper in which observer selection was examined. The present paper will go into even more detail on individual differences on several different performance measures and on combinations of measures.

The present study was done to obtain data and equations relating various observer performance measures to aircraft speed. Secondary purposes were to examine in detail both the nature of nontarget objects mistaken for targets and the types and amounts of individual differences in the ability of radar observers.

The ability of trained SLR radar observers to find and recognize unbriefed targets of specified types at four different image motion rates on the display simulating aircraft speeds ranging from 700 to 2100 knots was investigated. A single strip of high-resolution SLR imagery was magnified and displayed on a 14 x 14 inch screen. The observers were 20 SAC and TAC radar navigators whose task was to find all unbriefed airfields, dams, industry, railroad yards and tank farms whose images appeared on the display. Unlike the 1964 Rhodes and Self study mentioned earlier, four aircraft speeds instead of two were used to permit curve-fitting equations to be derived. In addition, the images of all false positives were photographed and recorded and false positive images were carefully examined.

EXPLANATION OF TERMS

Accuracy: The ratio of number of targets detected to the sum of detections and false positives. Thus, it is the proportion of responses that are detections, or the probability that what is identified as a target is a target. It is sometimes given as a percentage by multiplying the proportion by 100.

Agreement Coefficient: An index of response similarity, i.e., a number that indicates the similarity of an individual's choices to the choices of the experimental group of which he is a member. It is computed by counting, for each object that he selects, the number of people who select the same object, summing over all of the objects that he selects and dividing by the product of the number of people in the group and the number of objects that he selects. Its size can vary from 0 to 1.

Aircraft Speed (or Simulated Aircraft Speed): The speed over the terrain corresponding to the rate of motion of the displayed image across the viewing screen. Numerically, it is image motion rate times the reciprocal of image scale.

Analysis of Variance: A powerful statistical procedure or technique for the analysis of data which is used when more than two means (or conditions) are to be compared to determine if obtained differences are significantly different (genuinely different) or whether the differences are attributable to chance, i.e., not likely to be obtained upon repetition of an experiment. Its name comes from its use of variabilities and their ratios.

Completeness: The following definitions are synonymous: (1) the proportion of targets that are detected, (2) the ratio of the number of targets detected to the number present, and (3) the average probability of detection. Completeness, when multiplied by 100, becomes percentage of targets detected.

Confidence (or Subject Confidence or Confidence Level): The observer's certainty that the object he has designated as a target is indeed a target of the type that he has indicated. The observer indicates confidence level by depressing the appropriate switch.

Detection: A response by an observer or test operator indicating to the test administrator that a target is present. For example, a target is designated by the observer's hand-held stylus, as evidenced in a picture from the data camera which shows both the display and the stylus. This is an operational definition and example. From the subjective point of view of the observer or test operator, detection takes place when it is decided that an area on the displayed image represents a target. Often, detection is not distinguishable from recognition by the test administrator or even by the observer.

Efficiency: A measure or index of excellence of performance that takes into account more than one important characteristic of task execution. In this report, operator efficiency is defined as the product of accuracy and completeness.

F: In tables of analysis of variance "F" is the ratio of two of the variances which appear in the table. By referring to statistical tables one may find the probability of obtaining an "F" as large as or larger than the obtained "F" by chance alone.

False Positive: A nontarget object identified as a target by an observer: a portion of the displayed image is designated as a target when no target is present at the corresponding ground area. Some authors define a false positive as a recognition response made when a real target is not present. False positives are also referred to as false alarms, false targets, or spurious targets.

Ground Travel: The distance traveled over the terrain in the interval between the appearance on the display of an image of an object and its detection by an observer when the interval between receiving of a radar return and display upon a screen to the observer is practically instantaneous. If a processing delay takes place in the

radar equipment, then the distance covered by the aircraft in this time period would have to be included. See "Screen Travel."

Mean: An average or measure of central tendency. There are several sorts of mean, but in statistics, unless otherwise stated, the mean is the arithmetic mean, which is the sum of the scores (or measures) divided by the number of scores: M = (Sum X)/n.

Observer: A person being tested (see "Subject").

Prorating: Dividing unscorable observer responses (designations) between detections and false positives according to the proportion of detections and false positives occurring in the scorable responses. Unscorable responses may occur when the observer's head or arm gets between the data camera and the display.

Radar (Radio Direction and Ranging): A device that emits (or transmits) electromagnetic energy in the radio spectrum and utilizes the reflections of this energy from objects to obtain ranging and direction information. Imaging radars, such as SLR, have a pictorial display of information obtained from the radar return.

Radar Return: The image on the display of an object or area on the terrain that is characterized by a radar reflectivity different from that of the area immediately surrounding it. On the display the image is lighter or darker than the area surrounding it.

Rear Projection Display: A display in which the subject looks at the image formed on a translucent screen by an optical projector located behind the screen.

Recognition: A target is correctly classified, i.e., assigned to the proper category. Subjectively, a target is recognized when the observer attaches a name to it that distinguishes it from other types of target objects. A recognized target is always a detected target, although the converse may not be true.

Reconnaissance (Aerial): A survey conducted to obtain, by use of an airborne vehicle and a sensing device, information about an area. It is usually an exploratory military survey of the territory of a real or potential enemy, and the information may be about such things as geography, resources, the activity of men and industry, etc. Often the purpose is to locate and/or evaluate targets.

Response: The subject indicates, with his stylus, and by use of the pushbuttons on the console, that an area of the displayed image represents a target of a certain type. If he is correct, his response is a detection response (or simply a detection), and if he is wrong, the response is a false positive response.

Screen Travel (or Screen Position): The distance moved down the screen by the image of an object before the subject photographs it with the data camera. It is a measure of how quickly objects identified as targets are detected.

Sensor: A device, mechanism or organism whose behavior (or output) indicates the presence of and/or the nature of objects, materials or energies external to itself. Airborne reconnaissance sensors indicate (record or display) some of the characteristics of the environment by their response to energy emitted by or reflected from objects. Cameras, closed-circuit TV, and radar are imaging sensors utilizing electromagnetic energy.

Simulated Speed: See "Aircraft Speed".

Simulation: A situation, usually in a laboratory or test facility, in which some aspects of an operational or field situation are duplicated or imitated. For example, rate of motion of the image on a display may simulate that which would obtain in the case of an aircraft with an imaging sensor moving at a given speed over the terrain.

SLR: Side-Looking Radar. Sometimes called SLAR, the "A" standing for "Airborne". See discussion in the text.

Subject: An individual serving as a test operator in an experimental study. Subject, test subject, test operator, operator, and observer are synonymous terms.

Standard Deviation: A measure of variability or scatter about the average or arithmetic mean. In a normal or Gaussian distribution ± one standard deviation about the mean includes approximately 64% of the cases (or area) of the distribution.

"t": Student's "t," a statistical quantity calculated from the data to test for the "genuineness" of obtained differences between averages. By referring to standard "t" tables, the probability can be estimated of obtaining, by the working of chance (or sampling) alone, a difference as large as or larger than that found in observer testing when the real or true (or "population") difference is zero.

Tank Farm: An above ground group of storage tanks, usually for storing oil or gas.

Target of Opportunity: (1) An unbriefed target, (2) a target whose presence was not known to the observer before he examined its image on a display.

Variance: In statistics, the square of the standard deviation. Sometimes used to indicate the amount of variability or scatter about the mean.

SIDE-LOOKING RADAR

The side-looking radar (SLR) system of an aircraft views a strip of terrain parallel to the flight path of the aircraft, but lying off to one side. Unlike other types of radar, SLR "illuminates" objects and picks up their radar reflections only once, when the aircraft passes by them. The motion of the aircraft thus moves the radar beam along a strip of ground, in effect "sweeping" out the terrain strip.

The image of the terrain moves down the display screen at a rate proportional to the speed of the aircraft over the ground. The target images come into view at the top of the screen, move down the display, and move off of the display at the bottom of the screen. The geometry of the aircraft-terrain-target situation and the time factor are as shown in figure 1.

An image of terrain a few miles wide is recorded on film before it is displayed to a radar observer. New territory is being recorded while old territory is on display. The display presents a continuous nonrepetitive image of the terrain and, unlike aerial photographs, variation in image scale across the scene is negligible. The displayed image looks somewhat like a relief map. The SLR pictures used in this study were taken by a Goodyear Aerospace Corporation APS-73 (XH-3) SLR. It has a ground resolution of about 50 feet. This radar is currently (1977) being used in South America for geological and agricultural mapping. The nature of SLR imagery may be seen by examination of the radar images in figures 2 and 4. The images used in these figures were produced by a Westinghouse Electric Corporation APQ-56 (XAA) Side-Looking Radar. Consent to reproduce them was given by the Westinghouse Aerospace Division. The resolution of this radar, as seen in the examples, is similar to that viewed by the subjects. The contrast, scale, width of terrain displayed, and size of targets differ very little from that of the radar images displayed to our subjects. An example of SLR that is in the form of a two-page picture of San Diego Bay and the surrounding countryside may be seen in the Sept. 8, 1967, issue of *Aviation Week* and *Space Technology*. The front cover of the Oct. 1977 *Scientific American* has a SLR picture on it, and further examples are given in the accompanying article by Jensen, et al. (1977). The Jensen article is a good introduction to SLR for the layman. Readers who desire some technical details of SLR may wish to examine R. O. Harger's 1969 textbook on the theory and design of synthetic aperture radar systems. It does not show examples of radar imagery.

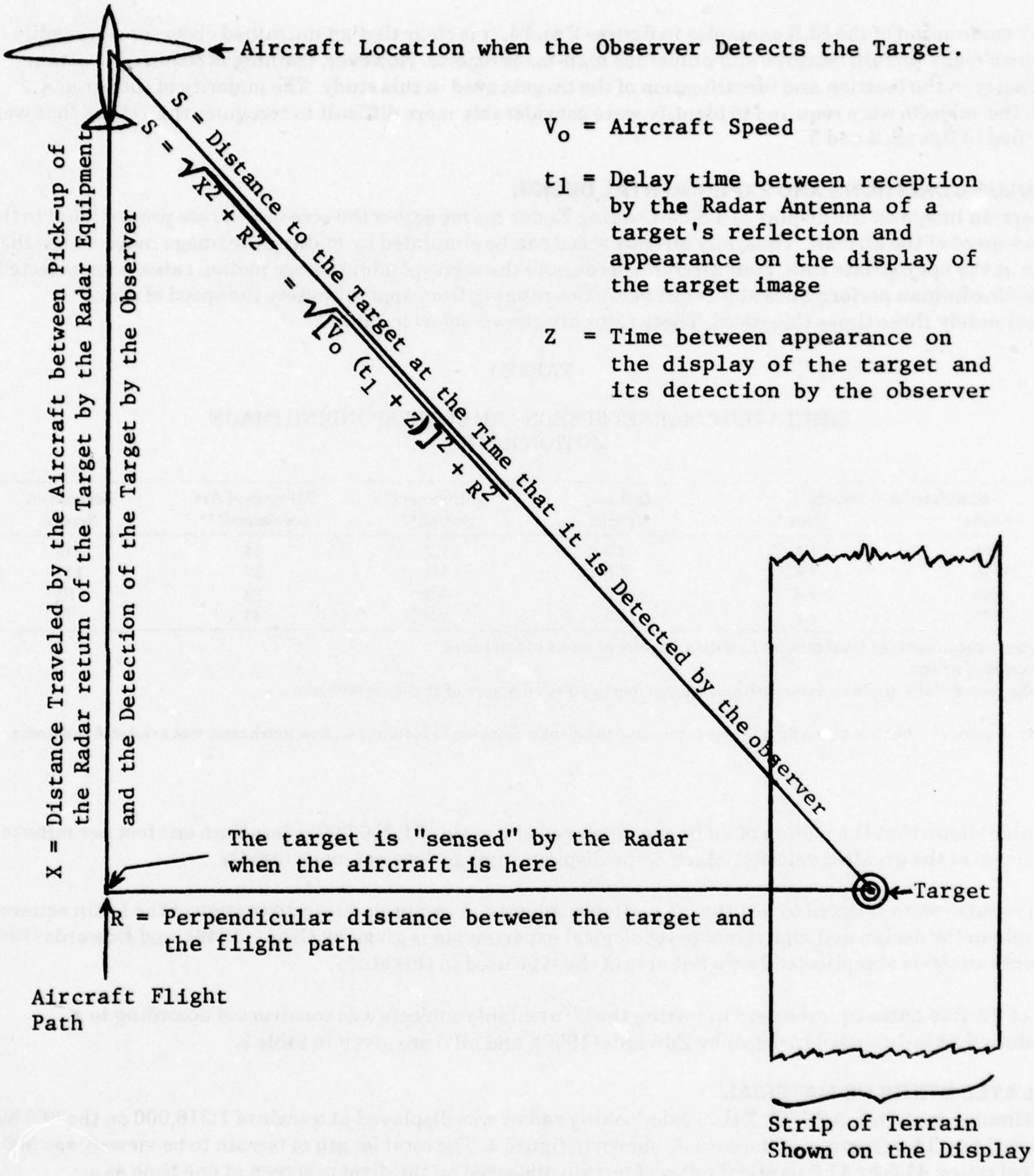


Figure 1. Time and Distance Relationships with a Side-Looking Radar.

From examination of the SLR examples in figures 2 and 4, it is clear that an untrained observer can readily recognize many terrain features and numerous man-made objects. However, training is required to attain proficiency in the location and identification of the targets used in this study. The majority of the targets which the subjects were required to identify were considerably more difficult to recognize than those that were identified in figures 2 and 3.

STIMULUS CONDITIONS AND EXPERIMENTAL DESIGN

The terrain image on the display of a Side-Looking Radar moves across the screen at a rate proportional to the ground speed of the aircraft. Thus, any aircraft speed can be simulated by making the image move across the screen at the appropriate rate. Four aircraft speeds, and the corresponding image motion rates, were selected to examine human performance at aircraft velocities ranging from approximately the speed of sound to approximately three times this speed. These rates are shown below in table 1.

TABLE 1

SIMULATED AIRCRAFT SPEEDS AND CORRESPONDING IMAGE MOTION RATES

Simulated A/C Speeds Knots	Mach*	Inches/ Minute	Millimeters/ Second**	Minutes of Arc per Second***	Seconds on Screen
700	1.1	3.9	1.7	14	215
1170	1.8	7.6	3.2	27	111
1640	2.5	9.2	3.9	33	91
2110	3.2	11.5	4.9	41	73

**Assuming standard sea level conditions with a velocity of sound of 658 knots.*

**At center of screen.

***At the center of the display screen with an assumed eye-to-screen distance of 16 inches (406 mm).

Seconds on screen is the time taken for a target to traverse the display from top to bottom, i.e., how much time was available to detect a target.

The table shows that the motion of an image displayed at a scale of 1:216,000 is less than one foot per minute. Thus, even at the greatest velocity, Mach 3, the displayed image does not move rapidly.

Latin squares were selected to test the 20 available subjects. A comprehensive discussion of the Latin square principle in the design and analysis of psychological experiments is given by Grant (1948), and Edwards (1950) discusses analysis of replicated Latin Squares of the type used in this study.

Each of the five Latin squares used in testing the 20 available subjects was constructed according to a procedure that is discussed in detail by Edwards (1950), and all 5 are given in table 2.

DISPLAYED STIMULUS MATERIAL

The stimulus material, APS-73 (XH-3) side-looking radar, was displayed at a scale of 1:216,000 on the 360 by 360mm (14 by 14 in.) screen of the console shown in figure 4. The total length of terrain to be viewed was 500 nautical miles; 41.5 by 41.5 nautical miles of terrain appeared on the display screen at one time as a continuously moving image. Parts of Kentucky, Tennessee and Oklahoma were covered on this strip.

Targets to be located and identified included 11 airfields, 8 dams, 19 industrial sites, 30 railroad yards and 10 tank farms. The frequency distribution for the occurrence of the 78 targets along the flight path is shown in figure 5. Various maps and charts were used to assist in locating targets on the film strip. For scoring observer responses, the image of an object known to represent a target was counted as a target only if it met this criterion: *In the opinion of the experimenter, the image, when stationary and pointed out to an observer, could be recognized occasionally.*

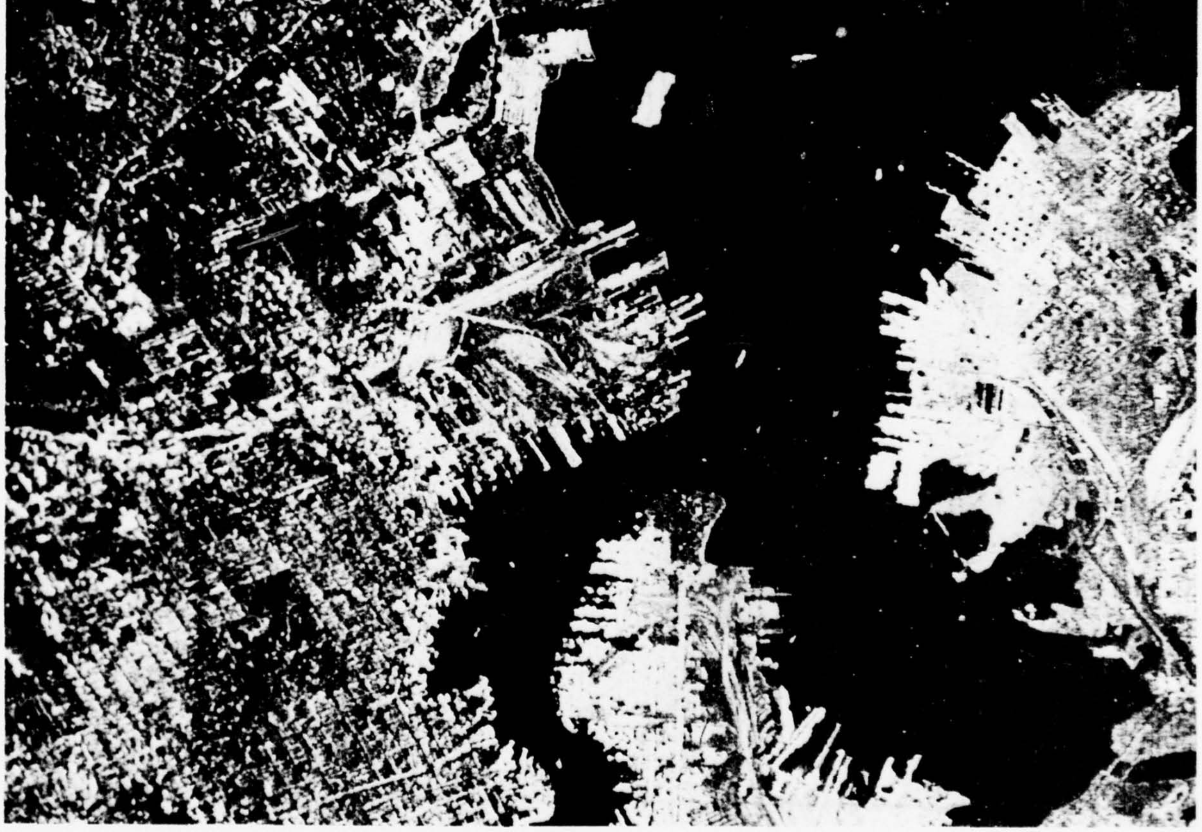


Figure 2. A picture of part of Baltimore Harbor made with a high-resolution side-looking radar.

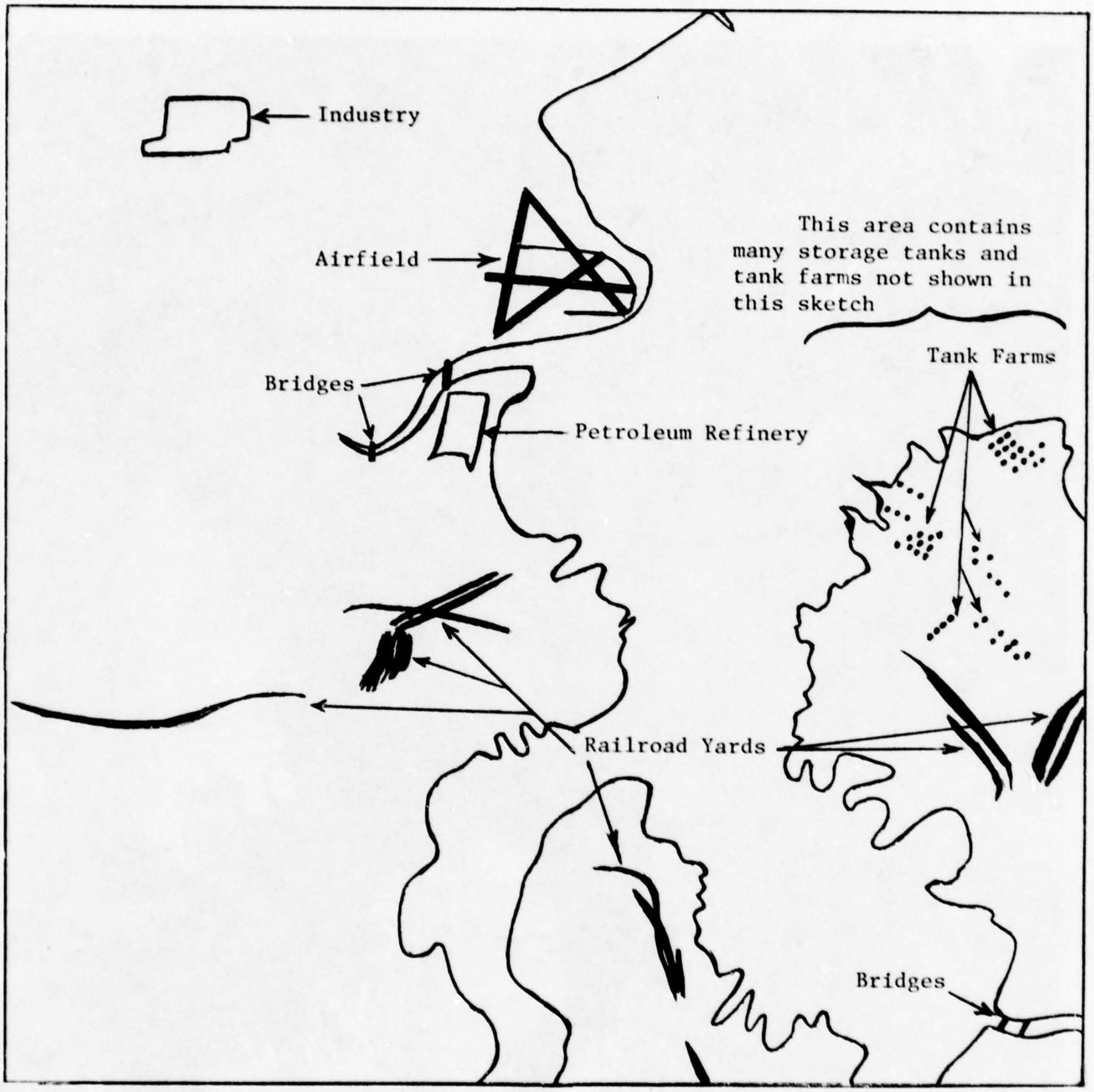


Figure 3. Key to some of the many objects contained in the SLR picture. Only a few of the more obvious objects are indicated on this page. Note the many streets, buildings and docks.

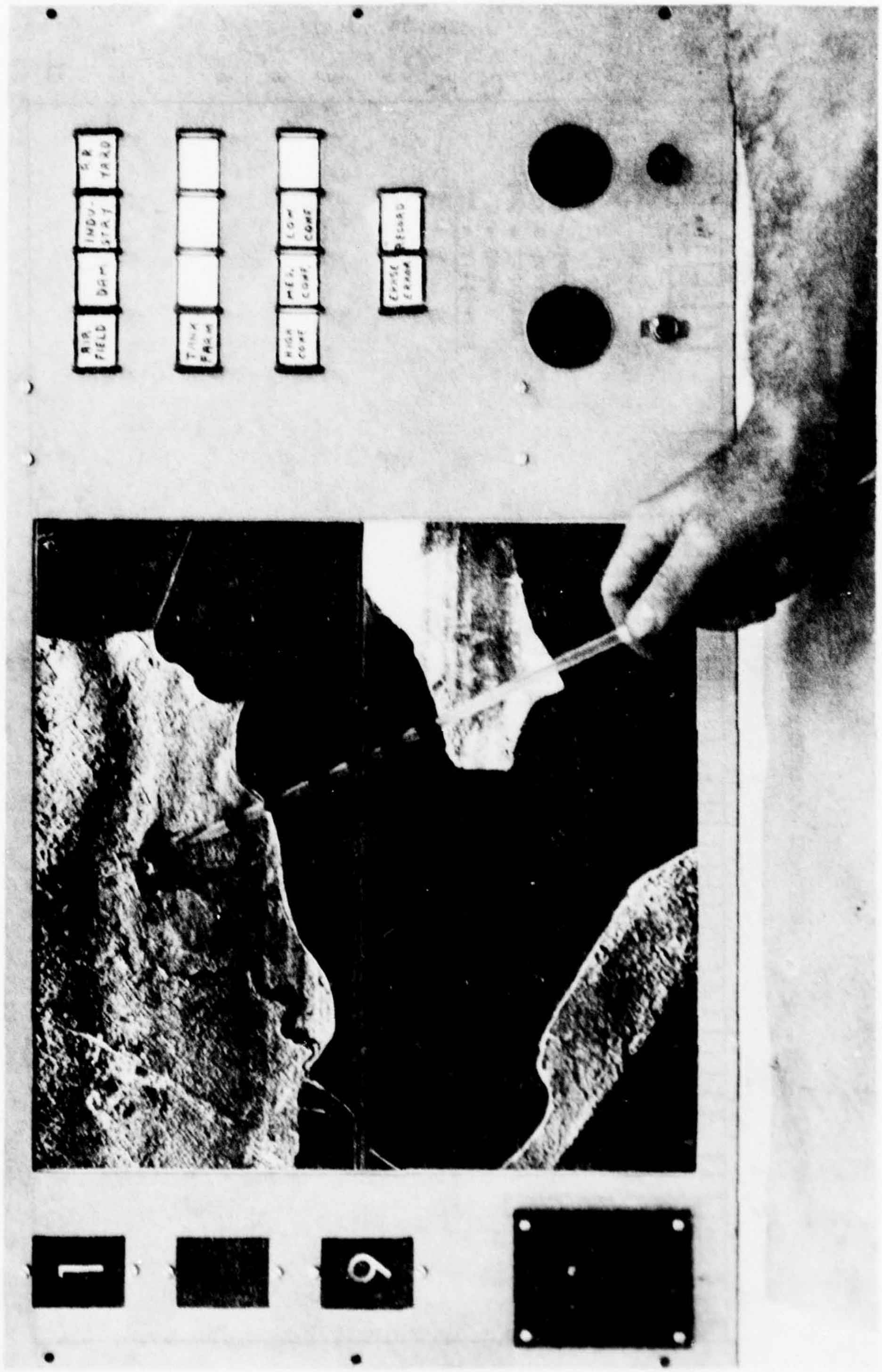


Figure 4. **A portion of the display console. The 1 in the top square on the left indicates that the subject has punched the airfield button, and the 9 indicates that he has reported a high confidence that it is an airfield. The picture on the display is not from the radar used in the present study.**

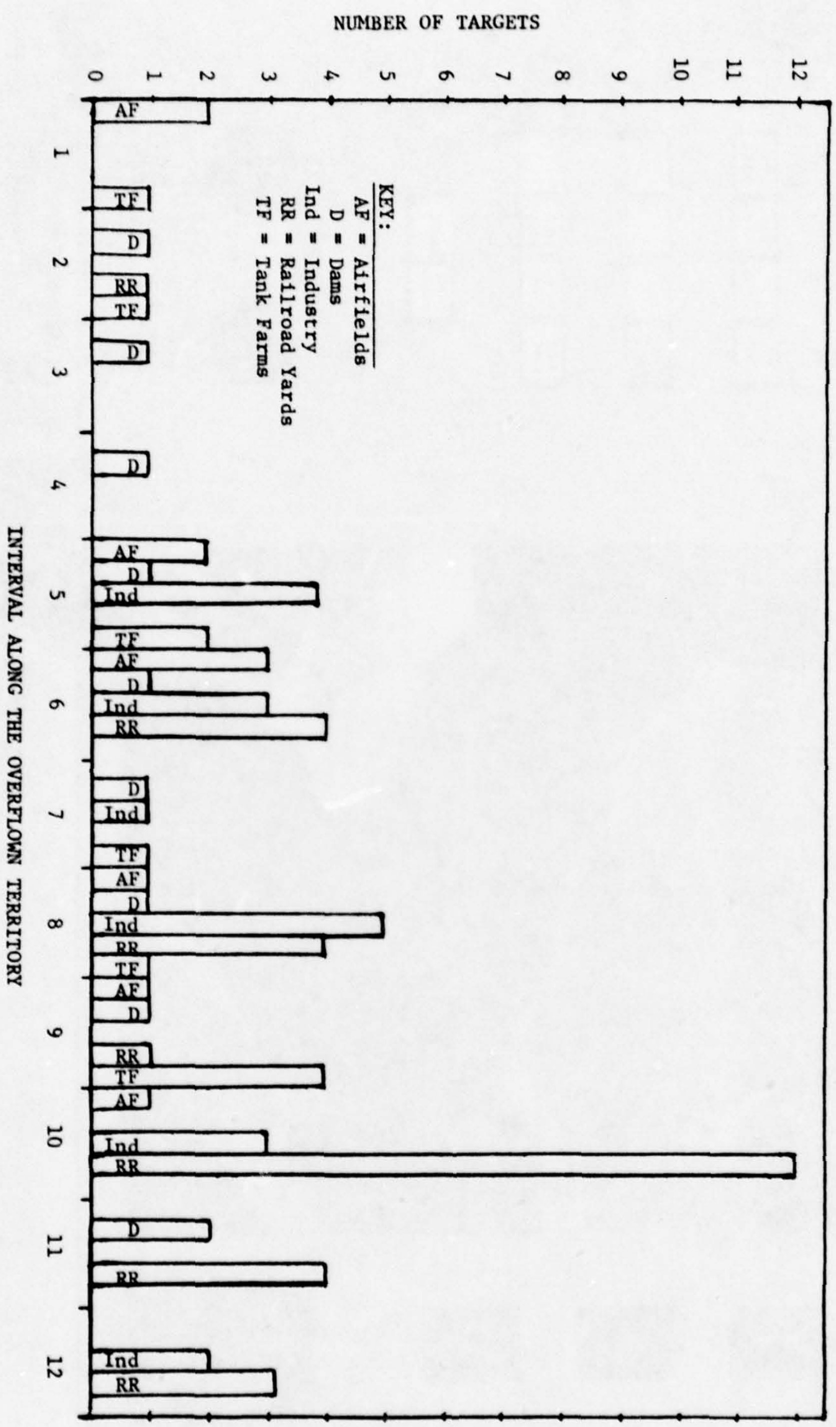


Figure 5. Distribution of targets along the terrain. Each of the 12 equally-large areas shown covers 41.5 by 41.5 nautical miles of terrain, corresponding to the height and width of the terrain.

TABLE 2

LATIN SQUARES

Latin Square	Subject	Trials 1	2	3	4
1	A	700	1640	1170	2110
	B	1170	700	2110	1640
	C	2110	1170	1640	700
	D	1640	2110	700	1170
2	E	1640	2110	700	1170
	F	2110	1640	1170	700
	G	1170	700	1640	2110
	H	700	1170	2110	1640
3	I	2110	1640	1170	700
	J	1170	700	2110	1640
	K	700	2110	1640	1170
	L	1640	1170	700	2110
4	M	2110	700	1640	1170
	N	1170	1640	2110	700
	O	700	2110	1170	1640
	P	1640	1170	700	2110
5	Q	1640	2110	700	1170
	R	700	1170	2110	1640
	S	2110	1640	1170	700
	T	1170	700	1640	2110

TEST SUBJECTS AND THEIR TRAINING

The 20 subjects were U.S. Air Force Radar Navigators from the Strategic and Tactical Air Commands (SAC and TAC). Except for a 10-hour training course, none of them had any previous training or experience with side-looking radar. Prior to participation in this study, they had received at the Aerospace Medical Research Laboratory a 10-hour training course prepared by the author and his co-workers, had been tested for about four hours in other ground-based experiments and for one and one-half hours in airborne experiments, and had received an additional hour of training. Thus, at the start of the present investigation they were thoroughly familiar with the task and procedures involved.

Subjects were told that the goal of the training was preparation for experimental conditions, not preparation for operational duties, and that the allotted number of training sessions was not sufficient to produce experts. Subjects were also told that no attempt had been made to simulate an operational mission.

The training program to familiarize subjects with side-looking radar displays included five sessions, each of two and one-half hours duration. Study of the side-looking radar imagery began with AMRL Memorandum P-64, *Side-Looking Radar Training Material for the 665A Program*. This training manual contained introductory material on (1) the physics and technology of side-looking radar, (2) a comparison of conventional bombing navigation radar and side-looking radar, (3) a discussion of imagery degradation, (4) the target signatures and target logic for targets of the types used in the studies, and (5) positive prints of SLR imagery containing relevant targets. Other training materials included side-looking radar film and corresponding transparent overlays showing the location of all targets on the film.

Continuous strips of imagery matching the overlays were available for projection on an optical viewer with a variable film speed control. Subjects viewed the imagery moving on the display device at varying speeds for approximately half of each training session and studied corresponding pictures on which all targets had been labeled. Training material was divided into training sets and subjects were tested after completion of each set.

The tests required identification of circled targets. At no time during training did the subjects view radar imagery of the particular terrain or targets used in the present study.

EXPERIMENTAL SESSIONS AND INSTRUCTIONS TO OBSERVERS

Observers were seated in a darkened room at the console shown in figure 4. An observer could move his chair or himself in the chair to obtain whatever viewing distance that he desired. The task was to find and to designate with an illuminated stylus, all of the targets within the five categories of targets used in the study.

The following instructions were read out loud to each observer:

> During this experiment you will be viewing a strip of positive, side-looking radar imagery. There is no briefing or briefing material. In other words, you will not know the location of the terrain imaged or the direction of the flight on which the imagery was collected. The width of the terrain covered is 41.5 nautical miles and will fill the full display screen. The simulated aircraft velocities will vary from 700 to 2110 knots. On this trial the simulated aircraft velocity is * knots. You will have four trials in all. The total experimental time will be about one hour and 15 minutes spread over a period of eight hours. You will be performing on other side-looking radar studies between trials on this particular film strip.
>
> The targets for this experiment are (1) airfields, (2) dams, (3) industry, (4) railroad yards, and (5) tank farms and petroleum refineries. When you have located a target, indicate your confidence level by pushing one of the three confidence level buttons. Confidence level for each detection is defined as the certainty you have that you are right, and is indicated by depressing the high, medium or low confidence switch.
>
> The same procedure is used for recording each response.
>
> 1. *Locate the target.*
> 2. Push the appropriate target name switch.
> 3. Push the appropriate confidence level switch.
> 4. Point to the target with the stylus and push the Data Insertion Switch to Record position and hold it briefly. Then, push the switch to Advance position and release. Continue the search for additional targets.
>
> Remember to keep your head out of the field-of-view of the data camera. Do you have any questions?

---

*The speed to be used in the upcoming trial was read out: it was different for each trial.

RESULTS

A. NUMBER OF DETECTED TARGETS

The number of targets detected is frequently used as a performance criterion in target detection tasks. Table 3 gives this data for individual subjects at the four aircraft speeds simulated in the present study.

Large individual differences in number of targets detected are apparent in this table. At each of the four simulated aircraft speeds the subject detecting the largest number of targets found more than twice as many targets as the subject detecting the smallest number. Note, however, that the best and worst performance is not always from the same two people. When performance is averaged for the four speeds, the most efficient subject's 26.9 detections bears this same 2:1 relationship to the poorest subject's mean of 12.5.

The performance variability between observers and within observers is readily apparent in figure 6 which plots the data for individual observers at all four aircraft speeds. In the figure observers have been arranged along the horizontal axis in order of decreasing average performance. The table and the graph clearly demonstrate that from one test session to the next subjects vary considerably in performance, even when an allowance is made for differences in difficulty attributable to differences in simulated aircraft speeds. Also, it is clear from the repeated testing that some of the subjects are considerably more efficient at finding targets than are others.

A major concern of the present investigation was to compare the number of targets detected by an observer at various simulated aircraft speeds. An examination of the means at the bottom of table 3 and the plot of these numbers in figure 7 reveals a constant decrease in number of detections, from a high of 19.3 at 700 knots to a low of 16.2 at 2110 knots. A straight line may be fitted to the 80 target detection scores made by the 20 subjects by using a least squares criterion. The best fitting linear equation for predicting number of expected detections Y, from simulated aircraft speed, V, according to this criterion, is $Y = 20.828 - .002091\ V$. This is the line shown on the graph. Note how closely a linear function fits the data. The product moment correlation coefficient between the numbers of targets detected by observers and the numbers predicted by the equation is .9773. The probability of obtaining such a large r by chance alone is less than .05. The proper test for goodness of fit of the data to the prediction equation, however, is chi-square, which has a value of only .031. The probability of obtaining a value of chi-square at least as large as this by chance alone is over .99. Clearly, the fit of the prediction equation to the data is excellent. From an inspection of the graph, it is apparent that number of detections decreases slowly as aircraft speed increases. The number of targets detected at 2110 knots is 84.0% using the data points, and 85.2% using the best-fitting linear equation. Thus, a 300% increase in aircraft speed resulted in a 15% drop in number of targets detected.

At each aircraft speed the number of targets detected was different. A test may be applied to the data to determine if the obtained differences are artifacts attributable to sampl ng error. The procedure, known as analysis of variance, is described for independently drawn Latin squares by Edwards (1950). The hypothesis for this test is that there are no real differences in number of detections at different speeds.

The data (number of detections) to use in the test are obtained by counting rather than by measuring along a continuous distribution. Thus, its distribution is not Gaussian, and a square-root transformation of the data is required to make it suitable for analysis of variance. Use of this type of data transformation is discussed by Snedecor (1956), and also by Olds et al. (1956). Another requirement for analysis of variance is homogeneity of variance, which may be checked by a test devised by Bartlett and discussed for Latin squares by Edwards (1950). This test yielded a chi-square corrected for continuity of 7.34 with four degrees of freedom, a value which is not significant at the .05 level of significance. Thus, homogeneity of variance is an acceptable hypothesis.

Since the hypothesis of homogeneity of the error variance is tenable, the error sum of squares from the separate Latin squares and their degrees of freedom may be pooled to obtain a common estimate of error variance. Since the choice of error term for testing the significance of the velocity effect depends upon the

TABLE 3

NUMBER+ OF TARGETS DETECTED BY INDIVIDUAL SUBJECTS

Subject	Aircraft Speed in Knots 700	1170	1640	2110	Means for all Speeds Combined Number	Percent	Standard Deviation	Rank
A	25.808	27.643	24.732	23.353	25.384	32.544	1.810	2
B	17.531	16.304	30.000**	26.234**	22.517	28.868	6.665	4
C	12.414	11.600	16.286	10.946	12.812	16.426	2.3929	19
D	30.032**	33.177**	20.045	24.170	26.856**	34.431**	5.878	1
E	24.000	23.000	18.961	16.000	20.490	26.269	3.702	7
F	17.000	15.000	18.000	11.000	15.250	19.551	3.096	15
G	14.000	9.000*	14.000	13.000	12.500*	16.026*	2.381	20
H	15.000	20.870	17.362	18.000	17.808	22.831	2.415	8
I	15.000	20.250	17.000	11.000	15.813	20.273	3.870	11
J	21.000	19.760	25.000	21.000	21.690	27.808	2.283	5
K	10.244*	17.000	23.000	13.000	15.811	20.271	5.802	12
L	30.000	17.000	16.000	23.000	21.500	27.564	6.455	6
M	29.000	21.842	23.377	19.576	23.449	30.063	4.017	3
N	16.000	19.312	14.362	12.545	15.555*	19.942	2.875	13
O	29.000	19.679	9.118*	11.000	17.199	22.050	9.113	10
P	19.268	15.000	12.000	15.000	15.317	19.637	2.990	14
Q	16.000	13.709	17.425	13.567	15.175	19.455	1.869	16
R	12.000	17.000	12.000	14.280	13.820	17.718	2.375	17
S	18.000	18.383	16.000	18.000	17.596	22.559	1.079	9
T	14.000	12.000	17.000	9.000*	13.000	16.667	3.367	18
TOTAL	385.297	367.529	361.668	323.671	359.542	460.953	74.433	
MEAN	19.265	18.376	18.083	16.184	17.977	23.048	3.722	
S.D.	6.460	5.533	5.084	5.176	4.316			
RATIO++	2.93	3.69	3.29	2.91	2.15	2.15		

+Table entries are prorated (corrected) as explained in the text.
++Ratio or range ratio, is the ratio of the highest to the lowest score in the column.
*Lowest score in the column.
**Highest score in the column.

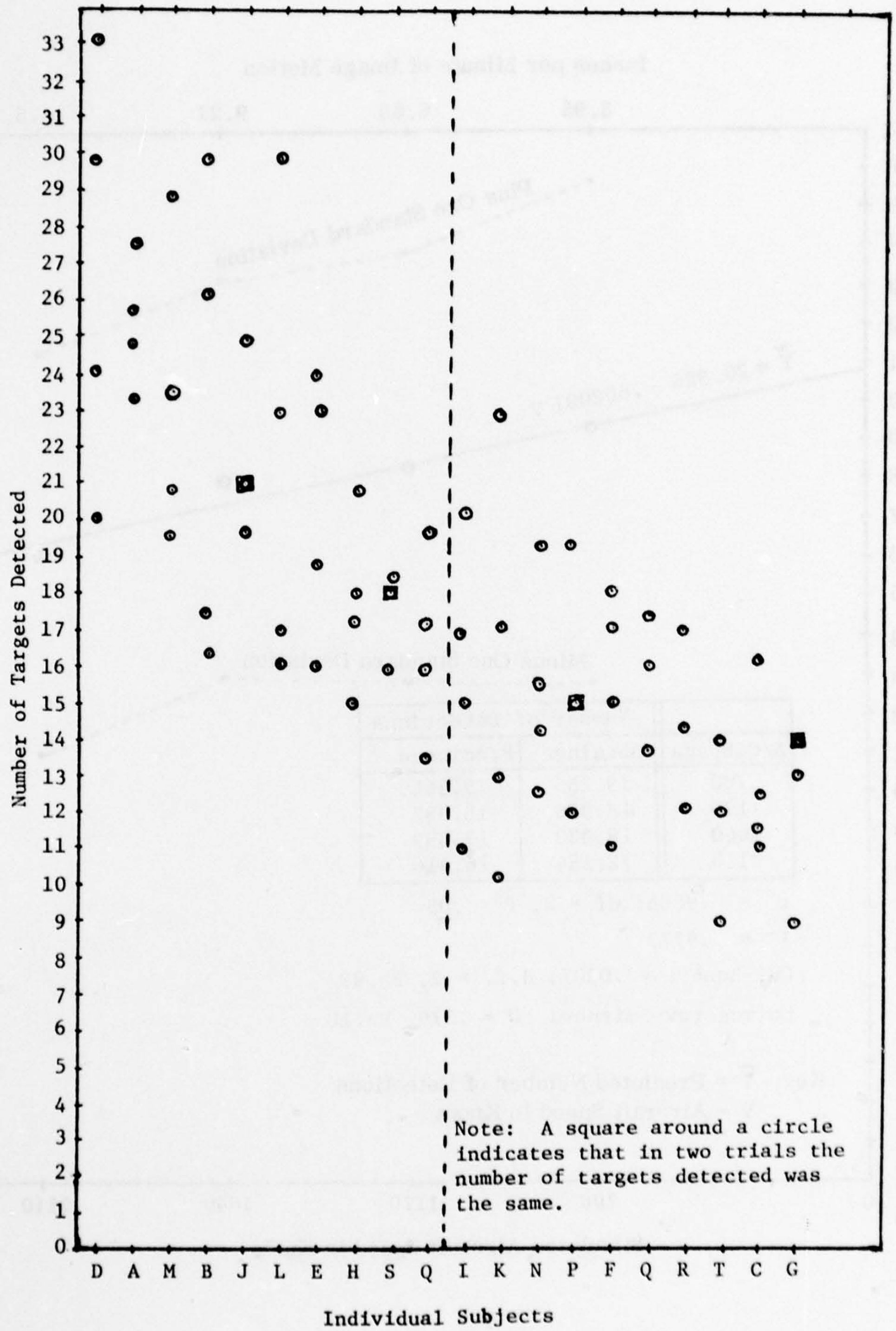
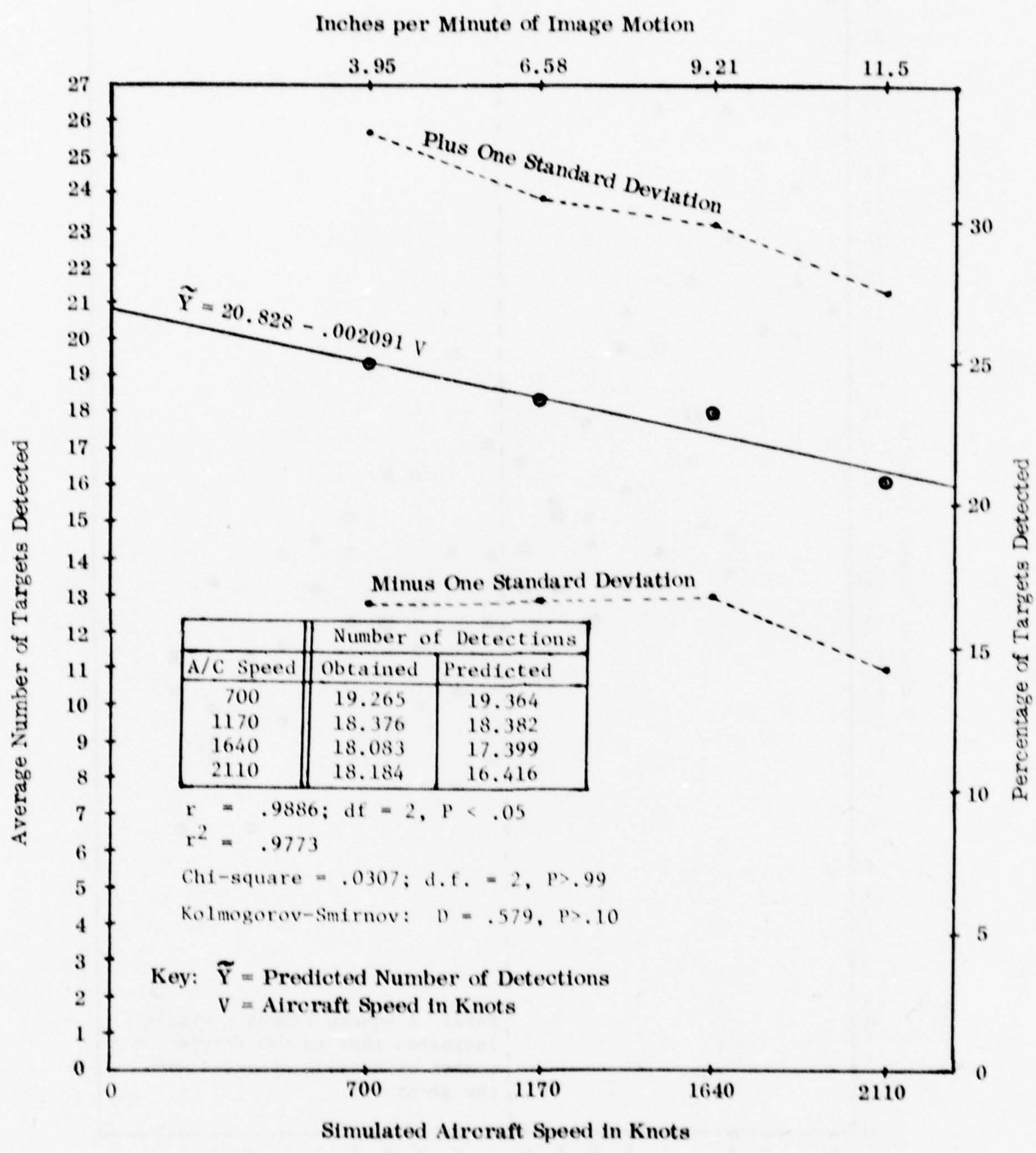


Figure 6. Number of targets detected by individual subjects at four different simulated aircraft speeds. Subjects are arranged in order of average performance.



	Number of Detections	
A/C Speed	Obtained	Predicted
700	19.265	19.364
1170	18.376	18.382
1640	18.083	17.399
2110	18.184	16.416

Figure 7. Number of targets detected at four different aircraft speeds. A least-squares best fit line is fitted to the circled data points (each of which is the mean for 20 subjects).

significance of the interactions, an analysis of variance for interactions is given in table 4. In this table neither the Latin squares by trials interaction nor the Latin squares by velocities interaction is statistically significant at the .05 level. The best common estimate of error which is obtainable from the data pools the error sums of squares from the separate Latin squares with the sum of squares for the interaction between Latin squares and velocities. Table 5 gives the analysis of variance. The most important result of the analysis is the finding that the number of detections at the different velocities are significantly different at the .05 level of statistical significance.

TABLE 4

ANALYSIS OF INTERACTIONS FOR NUMBER + OF DETECTIONS

Source of Variation	Sum of Squares	df	Mean Square	*F*
Latin Squares x Trials	3.5835	12	.2986	2.07
Latin Squares x Velocities	2.3665	12	.1972	1.36
Error	4.3391	30	.1446	

+Square root transformed data was used
NOTE: Neither F is statistically significant.

TABLE 5

NUMBER OF DETECTIONS +: ANALYSIS OF VARIANCE

Source of Variation	Sum of Squares	df	Mean Square	*F*
Latin Squares	5.3767	4	1.3442	8.42**
Between Subjects in Same Square	13.4854	15	.8990	5.63**
Velocities	1.4101	3	.4700	2.94*
Trials	2.6419	3	.8806	5.51**
Latin Squares x Trials	3.5853	12	.2986	
Error (including LS x Velocities)	6.7056	42	.1597	
TOTAL	33.2032	79		

+Square root transformed data was used for the analysis.
*, **Statistically significant at the .05 and .01 levels, respectively.

The foregoing parametric analysis, although more sensitive than a nonparametric test, requires certain assumptions that are not required by a non-parametric test. Thus, a Friedman two-way analysis of variance by ranks, discussed in detail by Siegel (1956), was used. This yielded a χ^2_r of 7.66 with an associated probability of over .05. Thus, the finding of the earlier and more sensitive test was not confirmed. In view of the small magnitude of the velocity effect upon number of detections and the large differences between different subjects and for the same subject when tested at different times, this is not a surprising finding. It does not invalidate the results of the parametric analysis which detected the performance decrease with increased aircraft speed.

Future experimental studies using small numbers of subjects to detect targets of the types used in the present investigation in the same range of aircraft speeds are not likely to obtain statistically significant differences in numbers of targets detected without using parametric statistical tests. Even then, the number of subjects may have to be increased to 10 or 20 per velocity: the large differences between and within subjects tends to obscure the small differences attributable to variation in aircraft speed.

The analysis of variance for individual Latin Squares, given in table 6, illustrates this point. Although each Latin Square contained four subjects, not one of the five squares yielded a statistically significant velocity effect upon the numbers of targets detected. Only by pooling the data of the five squares to obtain 20 subjects was the analysis made adequately sensitive to detect the influence of aircraft speed upon number of detections.

In the analysis of variance table, the mean square between subjects in the same Latin square is approximately twice as large as the mean square for velocities. The hypothesis of no difference between subjects in the same Latin square is rejected at the .01 level of significance: it is concluded that test subjects are significantly different in detection performance. Indeed, differences between individuals exceed differences attributable to aircraft speeds. From the above results, it appears that when the number of targets that an individual will detect must be predicted, knowledge of his detection capability in similar tasks is at least as essential as the aircraft speed that will be used.

In the table, the trial effect is significant at the .01 level of significance. The mean square for trials, .8806, is greater than that for velocities, .4700, by a factor of almost two. The means and standard deviations for number of detections at each trial are at the bottom of figure 8 which plots the data. In this figure a trend is seen for number of detections to increase with trials. As a point of interest, six comparisons could be made between the four trial means, by use of "*t*" tests. When this was done, in only one comparison, that between the

TABLE 6

NUMBER OF DETECTIONS (SQUARE-ROOT TRANSFORMED): ANALYSES OF VARIANCE FOR INDIVIDUAL LATIN SQUARES

Latin Square	Source of Variation	Sum of Squares	df	Mean Square	*F*
1	Velocities	.0920	3	0.0307	.166
	Subjects	6.3063	3	2.1021	
	Trials	1.6795	3	0.5598	
	Error	1.1099	6	0.1849	
	Total	10.0157	15		
2	Velocities	.3348	3	0.1116	1.455
	Subjects	2.1743	3	0.7247	
	Trials	.8499	3	0.2833	
	Error	.4601	6	0.0767	
	Total	3.8191	15		
3	Velocities	3292	3	0.1097	.866
	Subjects	1.8763	3	0.6254	
	Trials	2.6419	3	0.8806	
	Error	.7603	6	0.1267	
	Total	5.6077	15		
4	Velocities	2.8591	3	0.9530	4.118
	Subjects	2.2843	3	0.7614	
	Trials	.6474	3	0.2158	
	Error	1.3883	6	0.2314	
	Total	7.1791	15		
5	Velocities	.1615	3	0.0538	.5203
	Subjects	.8442	3	0.2814	
	Trials	.4067	3	0.1256	
	Error	.6205	6	0.1034	
	Total	2.0329	15		
	Sum of Totals	28.6545	75		

NOTE: None of the velocity effects are statistically significant at the .05 level.

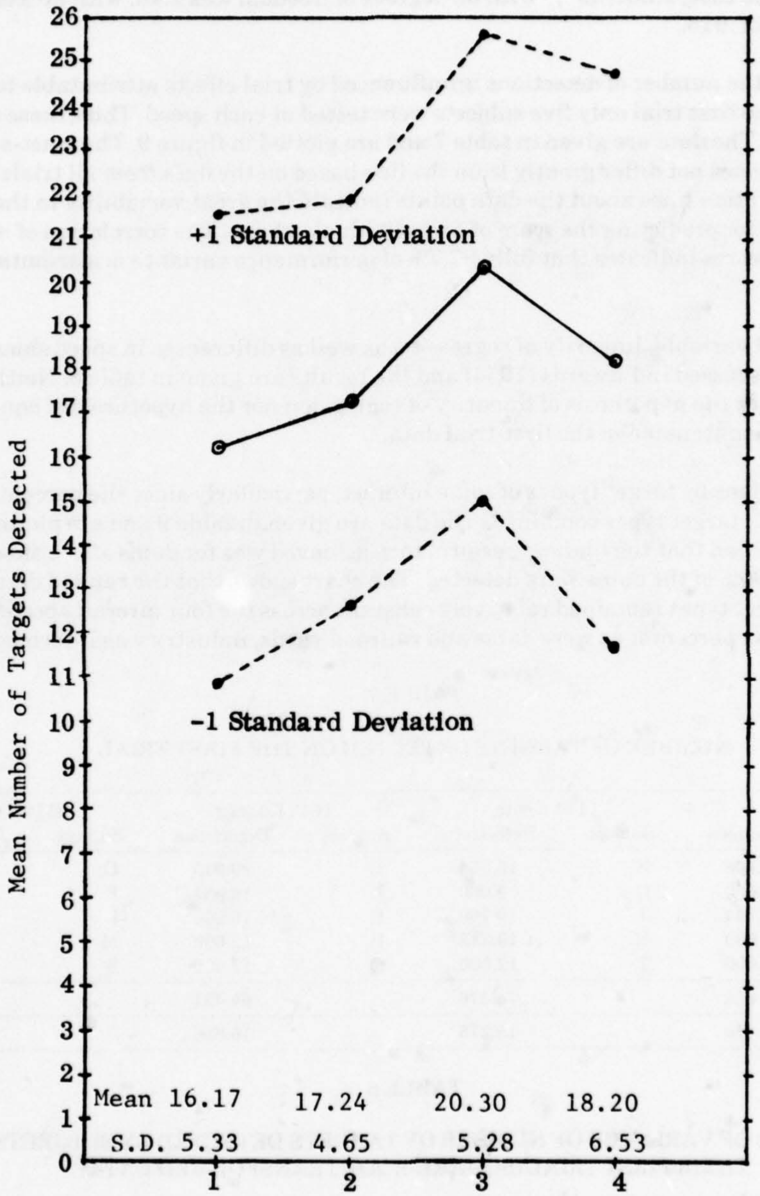


Figure 8. Number of targets detected in each of the four trials. Each of the four plotted points is the mean of twenty subjects, five subjects at each of the four velocities.

first and third trial means, was a difference between means found to be statistically significant at the .05 level of significance. In this one case, students "t" with 38 degrees of freedom was 2.46, with an associated probability of occurence of .015.

Only on the first trial is the number of detections uninfluenced by trial effects attributable to learning and/or boredom. However, on the first trial only five subjects were tested at each speed. Thus, these data are not likely to yield significant results. The data are given in table 7 and are plotted in figure 9. The least-squares best-fitting line shown on the graph does not differ greatly from the line based on the data from all trials. The plus one and minus one standard deviation lines about the data points indicate the great variability in the data and the low accuracy of the equation for predicting the score of an individual subject. The correlation of +.9886 between observed and predicted scores indicates that fully 97.7% of performance variance is attributable to aircraft speed.

Since speed is an ordered variable, linearity of regression as well as differences in speed should be examined. The tests for these are discussed in Edwards (1954) and the results are given in table 8. Neither "F" is significant, so that neither the hypothesis of linearity of regression nor the hypothesis of equal numbers of detections are found to be untenable in the first-trial data.

The breakdown of detections by target type is of some interest, particularly since the percentage of the targets detected was so low for all target types combined. The data are given in table 9 and are plotted in figure 10. From the chart it can be seen that the highest performance achieved was for dams at the slowest aircraft speed. Here, only about 30% of the dams were detected. The chart shows that the ranked detection percentages for the five different target types remained relatively constant across the four aircraft speeds. The target types with the highest detection percentages were dams and railroad yards, industry was intermediate, and

TABLE 7

NUMBER OF TARGETS DETECTED ON THE FIRST TRIAL

700 Knots		1170 Knots		1640 Knots		2110 Knots	
Subject	Detections	Subject	Detections	Subject	Detections	Subject	Detections
A	25.808	B	16.304	D	20.045	C	10.946
H	15.000	G	9.000	E	18.961	F	11.000
K	10.244	J	19.760	L	16.000	I	11.000
O	29.000	N	19.312	P	12.000	M	19.576
R	12.000	T	12.000	Q	17.425	S	18.000
Sum	92.052		76.376		84.431		70.522
Mean	18.410		15.275		16.886		14.104

TABLE 8

ANALYSIS OF VARIANCE OF NUMBER OF TARGETS DETECTED BY SUBJECTS ON THEIR FIRST TRIAL (SQUARE-ROOT TRANSFORMED DATA)

Source of Variation	Sum of Squares	df	Mean Square	F
Linear Regression	.4230	1	.4230	.934
Deviation from Regression	.9061	2	.4530	
Between Velocities	1.329	3	.433	1.05
Within Velocities	6.744	16	.422	
TOTAL		19		

NOTE: Neither F is significant at the .05 level.

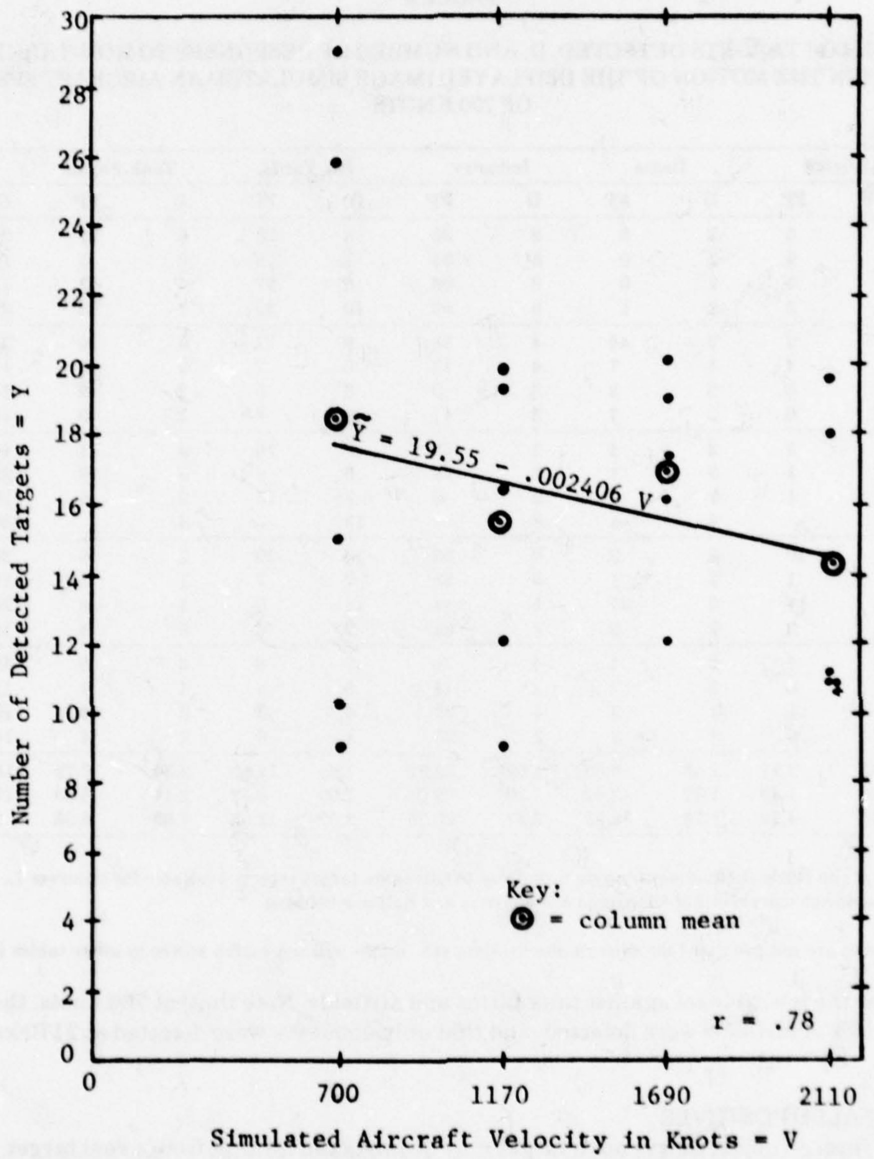


Figure 9. Number of targets detected by individuals on their first test run plotted against simulated aircraft velocity.

TABLE 9

NUMBER OF TARGETS DETECTED, D, AND NUMBER OF RESPONSES TO NON-TARGETS, FP, WHEN THE MOTION OF THE DISPLAYED IMAGE SIMULATED AN AIRCRAFT SPEED OF 700 KNOTS

Target	Airfields		Dams		Industry		RR Yards		Tank Farms		Totals	
Observer	D	FP	D	FP	D	FP	D	FP	D	FP	D	FP
A	1	0	2	5	9	25	6	26	6	17	20	73
B	1	0	2	0	5	34	7	8	2	5	17	47
C	1	0	2	0	2	66	6	37	1	2	12	105
D	5	5	2	1	8	83	10	42	2	26	27	157
E	3	3	3	45	4	58	9	27	5	14	24	147
F	1	4	4	7	4	13	6	7	2	6	17	37
G	1	0	2	3	3	9	6	5	2	9	14	26
H	0	0	2	7	3	17	8	8	2	9	15	41
I	0	3	2	1	1	36	8	16	4	1	15	57
J	2	1	3	1	7	29	6	8	3	0	21	39
K	1	1	2	4	2	8	3	13	2	5	10	31
L*	1	—	4	—	8	—	13	—	4	—	30	—
M	4	10	2	2	8	55	13	29	2	6	29	102
N	1	1	3	1	5	38	7	7	1	1	17	48
O	3	17	3	37	5	44	14	9	1	6	26	113
P	4	3	1	3	4	34	7	4	3	8	19	52
Q	0	1	2	1	4	20	7	8	3	5	16	35
R	2	6	2	2	2	18	5	4	1	1	12	31
S	1	1	3	3	4	23	8	2	2	4	18	33
T	3	5	3	3	2	23	4	6	2	4	14	41
MEAN	1.75	3.21	2.45	6.63	4.50	33.32	7.65	14.00	2.30	6.79	18.65	63.95
MEDIAN**	1.28	1.40	1.32	2.62	4.10	29.0	7.00	8.12	2.17	5.33	17.17	47.0
S.D.	1.45	4.32	.76	12.35	2.37	20.25	2.92	12.05	1.69	6.38	5.82	40.63

*Due to malfunction of the readout equipment, no data on false positives by target type is available for observer L.

**The median is the midmost score: Half of the scores are above it and half are below it.

NOTE: The above scores are not pro-rated for camera obscuration, etc., hence will not match scores in other tables in this report.

performance was at the lowest level against tank farms and airfields. Note that, at 700 knots, the slowest speed, only about 16% of airfields were detected, and that only about 9% were detected at 2110 knots, the highest speed.

B. NUMBER OF FALSE POSITIVES

A radar reflection from an object that is not a target may be mistaken for that from a real target. Such a nontarget object is frequently referred to as a false positive. It is also sometimes called a false target, a spurious target, or a false alarm. Ideally, an observer would not mistake any nontargets for targets, but in reality observers often make such mistakes. Obviously, the military utility of any airborne system in which observers make a larger number of such mistakes will be limited.

The number of false positives for the 20 subjects at each of the four aircraft speeds and for the average of the four speeds is given in table 10, and is shown in graphical form in figure 11. Individual subjects are arranged from left to right along the horizontal axis in order of increasing average number of false positives for the four aircraft speeds. The solid line on the graph represents the average number. From the graph it may be seen that individuals vary from a low of 24 (subject G) to a high of 130 (subject D). This is a ratio of 5:1. At each of the four aircraft speeds, the lowest-scoring subject reported approximately one fifth (or less) as many false

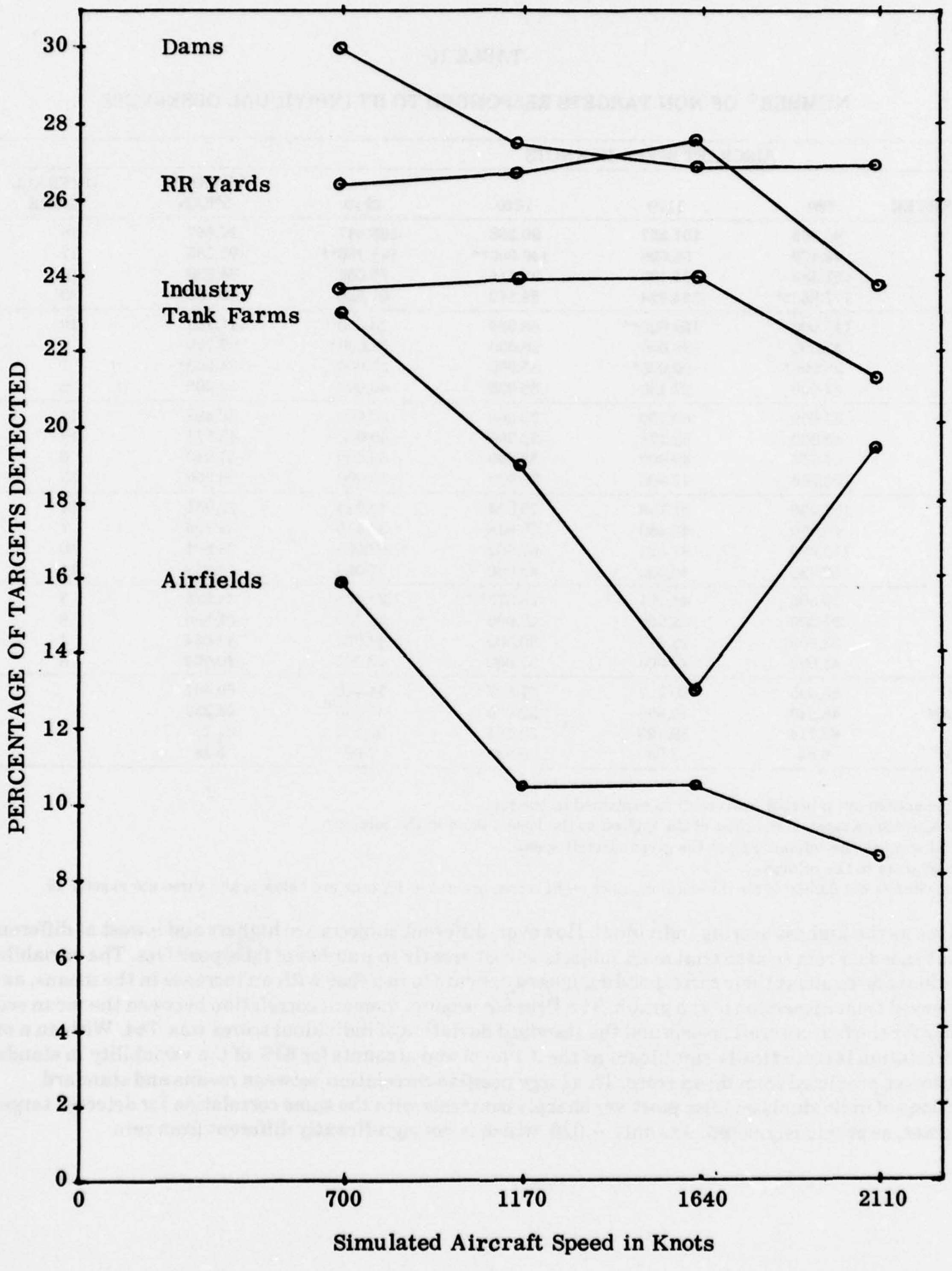


Figure 10. Percentage of Targets Detected by Target Types at Four Different Simulated Aircraft Speeds

TABLE 10

NUMBER+ OF NON TARGETS RESPONDED TO BY INDIVIDUAL OBSERVERS

	AIRCRAFT SPEED IN KNOTS					
OBSERVER	700	1170	1640	2110	OVERALL MEAN	OVERALL RANK
A	94.195	101.357	90.268	105.647	97.867	18
B	48.479	58.696	140.000**	143.766**	97.735	17
C	107.586	104.400	97.714	70.054	94.939	16
D	177.968**	154.824	88.163	97.830	129.696**	20
E	147.000	159.000**	68.039	73.000	111.760	19
F	37.000	38.000	26.000	18.000*	29.750	2
G	26.000*	20.000*	25.000	27.000	24.500*	1
H	41.000	27.130	30.638	40.000	34.692	5
I	57.000	60.750	76.000	32.000	56.438	12
J	40.000	32.292	62.000	40.000	43.573	10
K	31.756	49.000	38.000	33.000	37.939	6
L	106.000	43.000	31.000	93.000	68.250	13
M	102.000	61.158	76.623	48.424	72.051	14
N	48.000	42.689	32.638	33.455	39.196	7
O	113.000	67.321	67.882	67.000	78.801	15
P	52.732	48.000	42.000	37.000	44.933	11
Q	35.000	44.291	23.575*	23.433	31.575	3
R	31.000	52.000	47.000	36.720	41.680	9
S	33.000	29.617	38.000	30.000	32.654	4
T	41.000	22.000	57.000	40.000	40.000	8
MEAN	68.485	60.777	57.887	54.466	60.401	
MEDIAN	48.240	48.500	52.000	40.000≠	44.253	
S.D.	43.714	39.725	30.551	33.277	31.38	
RATIO++	6.84	7.95	5.93	7.99	5.29	

+Table entries are prorated (corrected) as explained in the text.
++Ratio, or range ratio, is the ratio of the highest to the lowest score in the column.
*Lowest score in the column, i.e., at the given aircraft speed.
**Highest score in the column.
≠The median is not definable for this column, since eight scores are above 40, nine are below it and three are exactly 40.

positives as the highest-scoring individual. However, different subjects are highest and lowest at different aircraft speeds. From trial to trial most subjects varied greatly in number of false positives. The variability of individual scores about their corresponding means appears to increase with an increase in the means, as may be observed from inspection of the graph. The Pearson product moment correlation between the mean score of subjects for the four aircraft speeds and the standard deviation of individual scores was .784. With an n of 20, the correlation is statistically significant at the .01 level and accounts for 61% of the variability in standard deviation as predicted from mean score. This large positive correlation between means and standard deviations of individuals on false positives sharply contrasts with the same correlation for detected targets. The latter, as previously noted, was only −.020, which is not significantly different from zero.

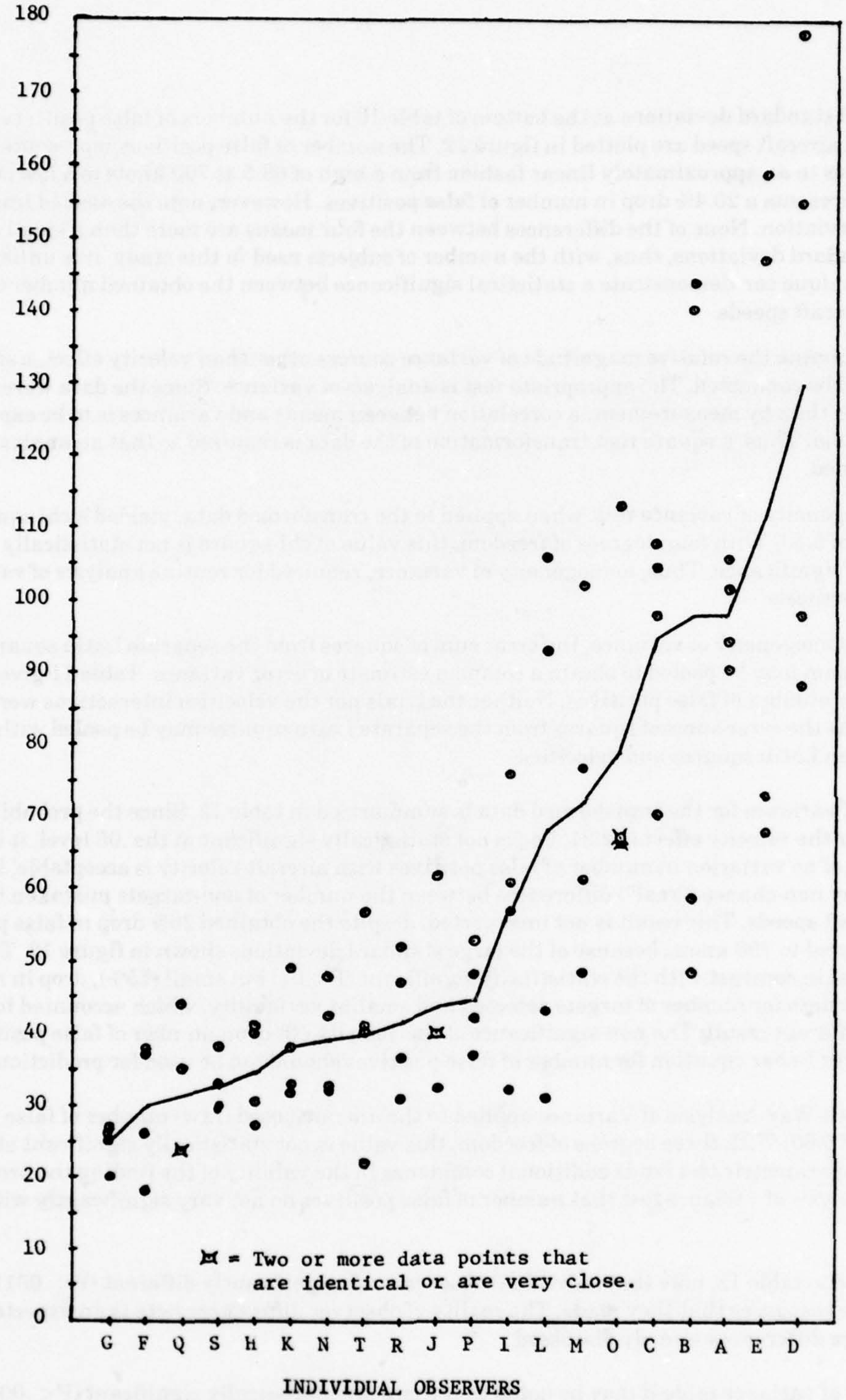


Figure 11. Number of False Positive Responses made by Individual Observers. Observers are arranged in order of increasing number of responses.

The means and standard deviations at the bottom of table 10 for the numbers of false positives at each of the four simulated aircraft speed are plotted in figure 12. The number of false positives represented by the four means decreases in an approximately linear fashion from a high of 68.5 at 700 knots to a low of 54.5 at 2110 knots. This represents a 20.4% drop in number of false positives. However, note the dashed lines representing ±1 standard deviation. None of the differences between the four means are more than a small part of any one of the four standard deviations, thus, with the number of subjects used in this study, it is unlikely that any statistical technique can demonstrate a statistical significance between the obtained number of false positives at different aircraft speeds.

However, to examine the relative magnitude of variance sources other than velocity effect, a statistical test of the data should be conducted. The appropriate test is analysis of variance. Since the data were obtained by counting rather than by measurement, a correlation between means and variances is to be expected, and indeed, was found. Thus, a square root transformation of the data is required so that an analysis of variance may be performed.

Barlett's homogeneity of variance test, when applied to the transformed data, yielded a chi square corrected for continuity of 5.33. With four degrees of freedom, this value of chi-square is not statistically significant at the .05 level of significance. Thus, homogeneity of variance, required for routine analysis of variance, is an acceptable hypothesis.

Since there is homogeneity of variance, the error sum of squares from the separate Latin squares and their degrees of freedom may be pooled to obtain a common estimate of error variance. Table 11 gives the analysis of interactions for number of false positives. Neither the trials nor the velocities interactions were statistically significant, thus the error sums of squares from the separate Latin squares may be pooled with the sum of squares between Latin squares and velocities.

The analysis of variance for the transformed data is summarized in table 12. Since the probability associated with the "F" for the velocity effect is 2.21, i.e., is not statistically significant at the .05 level, it is concluded that the hypothesis of no variation in number of false positives with aircraft velocity is acceptable, i.e., the data do not indicate any non-chance ("real") differences between the number of non-targets mistaken for targets at different aircraft speeds. This result is not unexpected, despite the obtained 20% drop in false positives at 2110 knots as compared to 700 knots, because of the large standard deviations shown in figure 12. This result for false positives is in contrast with the statistically significant ($P<.01$) but small (15%), drop in number of detections. The data for number of targets detected had smaller variability, which accounted for the statistically-different result. The non-significance of the velocity effect on number of false positives means that the least-squares linear equation for number of false positives should not be used for prediction purposes.

A Friedman Two-Way Analysis of Variance applied to the untransposed (raw) number of false positives yielded a χ^2_r of 5.60. With three degrees of freedom, this value is not statistically significant at the .05 level. Thus, this non-parametric test lends additional confidence in the validity of the finding the previous parametric analysis of variance test that number of false positives do not vary significantly with aircraft speed.

Referring again to table 12, note that individual observers are significantly different ($P< .001$) in the number of false positive responses that they made. The reality of observer differences here is an expected result in view of the very large differences already discussed.

In the analysis of variance table it may be noted that there is a statistically significant ($P< .001$) trial or order effect: The average number of false positive responses were different in the four trials. The means and standard deviations for each of the four trials are depicted in figure 13. Inspection of the figure reveals that the trial effect is large. The average number of false positives on the first trial is 46.9 and by the fourth trial or test session it has increased to 70.0, an increase of 49%. Although more false positive responses were present on the second trial than on the first trial, the increase was not statistically significant. However, although the number of false positives on the last two trials did not differ significantly from each other, both trials had significantly ($P< .05$) more false positives than did trial one.

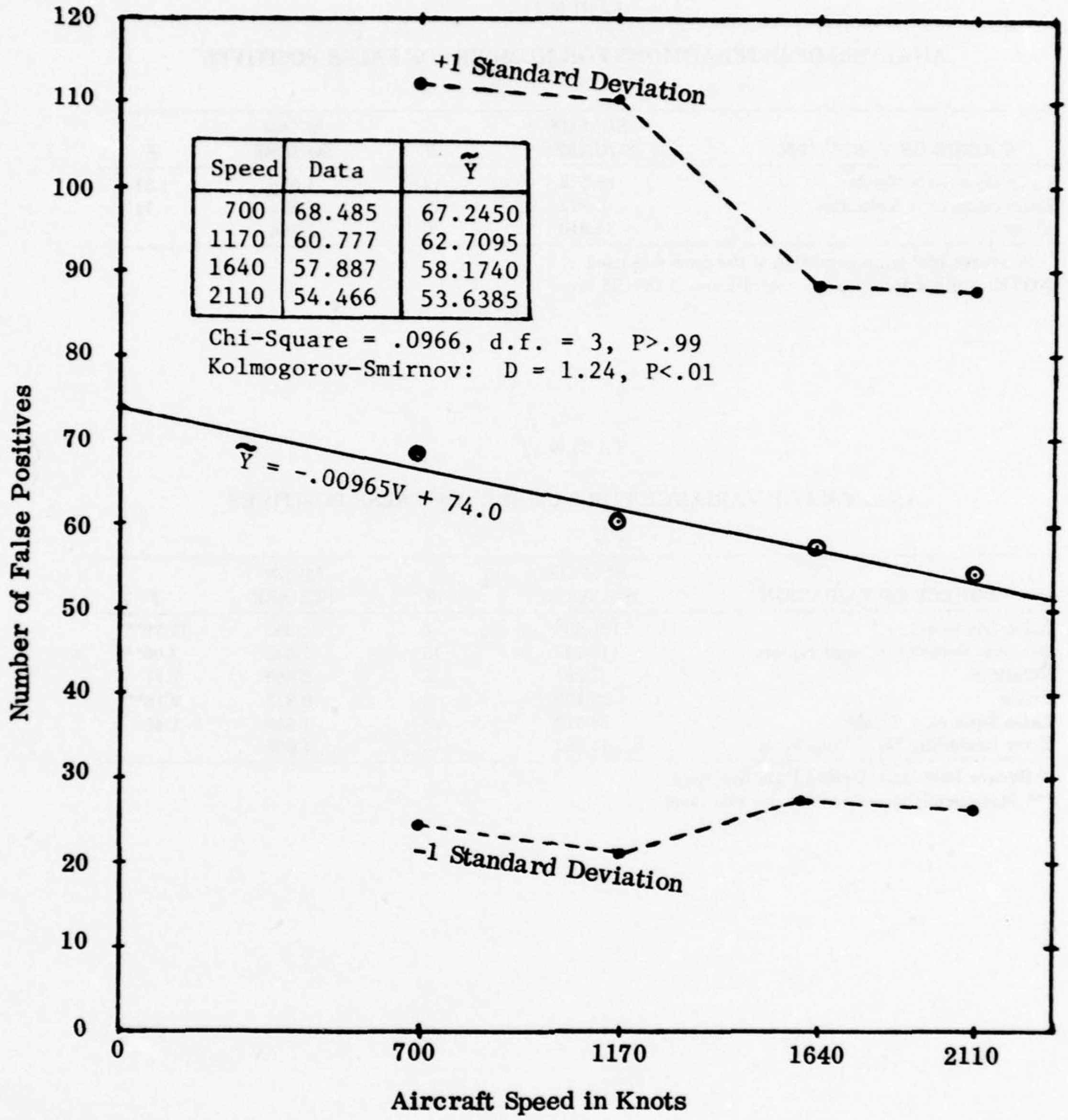


Figure 12. Number of false positives at various aircraft speeds. Due to the large variability, the velocity effect is not statistically significant, thus the graphed equation should not be used for predictive purposes.

TABLE 11

ANALYSIS OF INTERACTIONS FOR NUMBER+ OF FALSE POSITIVES

SOURCE OF VARIATION	SUM OF SQUARES	df	MEAN SQUARE	*F*
Latin Squares × Trials	18.572	12	1.548	1.33
Latin Squares × Velocities	9.972	12	.831	.71
Error	34.910	30	1.164	

+ A square root transformation of the data was used.
NOTE: Neither interaction is significant at the .05 level.

TABLE 12

ANALYSIS OF VARIANCE FOR NUMBER+ OF FALSE POSITIVES

SOURCE OF VARIATION	SUM OF SQUARES	df	MEAN SQUARE	*F*
Latin Squares	161.533	4	40.383	37.78***
Between Subjects in Same Square	113.147	15	7.543	7.06***
Velocities	7.081	3	2.360	2.21
Trials	29.439	3	9.813	9.18***
Latin Squares × Trials	18.572	12	1.548	1.45
Error (including LS × Velocities)	44.882	42	1.069	

+ Square Root Transformed Data was used.
*** Statistically significant at the .001 level.

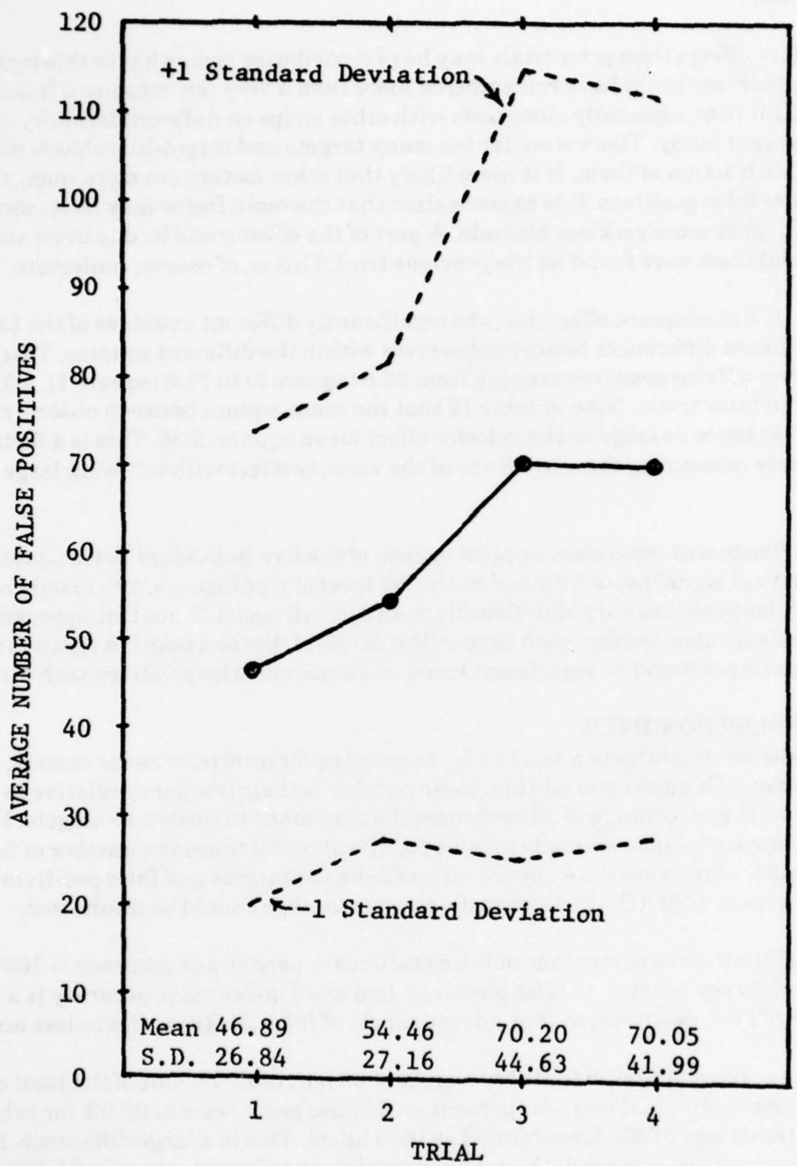


Figure 13. Average number of responses made to nontargets in each of the four trials or tests administered to observers.

Trial two was significantly worse ($P < .05$) than trial four, but was not significantly worse than trial three. In other words, observers were making more mistakes in the last two trials than on the first trial, but got no worse after the third trial.

It is possible that memory effects from prior trials may have contributed somewhat to this increase, but it appears unlikely that observers could have remembered more than a very few responses from earlier trials with the same strip of SLR film, especially since tests with other strips on different territory intervened between trials in the present study. There were far too many targets and target-like objects on the films to remember an appreciable fraction of them. It is more likely that other factors are more important in producing the increase in number of false positives. It is hypothesized that the main factor may be an increase in carelessness: A less critical or more reckless attitude. A part of the effect could be due to an attempt to find more targets on each trial than were found on the previous trial. This is, of course, conjecture.

The significant ($P < .001$) Latin square effect, i.e., the significantly different averages of the Latin squares, is most likely due to significant differences between observers within the different squares. This is supported by first-trial sums of number of false positives ranging from 26.6 (square 5) to 77.8 (square 1), a 3:1 ratio, with even larger differences in later trials. Note in table 12 that the mean square between observers in the same Latin square, 7.54, is over twice as large as the velocity effect mean square, 2.36. This is a further indication of the difficulty of accurately measuring the magnitude of the velocity effect without using large numbers of observers.

When an analysis of variance was separately applied to each of the five individual Latin squares, only in the fourth square was statistical significance attained at the .05 level of significance, i.e., in only one of five squares did number of false positives vary significantly with aircraft speed. Since this approach of individual square analyses involves repeated testing, each time at the .05 level, the one positive result is dubious. Indeed, the overall analysis of variance found no significant trend in number of false positives with aircraft speed.

C. PERCENTAGE OF FALSE POSITIVES

The number of false positives (nontargets mistaken for targets) is, for unbriefed radar targets, so large as to constitute a serious problem. Of more interest than their number is their frequency relative to all responses, conveniently expressed as the percentage of all responses that are made to these non-targets. Percentage of false positives (or percentage of responses made to false positives) is 100 times the number of false positives divided by the total number of responses, i.e., by the sum of detected targets and false positives. As a formula, percentage of false positives $= 100F/(D+F)$. Obviously, this percentage should be minimized.

It should be kept in mind that, since percentage of false positives + percentage accuracy $= 100F/(F+D) = 100D/(F+D) = 100$, % accuracy = 100 − % false positives. In a word, percentage accuracy is a linear transform of percentage of false positives, so that minimizing % of false positives maximizes accuracy.

The percentage of false positives for individual observers is shown in table 13. Out of the total eighty test runs or trials, the highest (worst) individual score for percentage of false positives was 90.0% for subject C at 1170 knots, while the lowest (best) was 57.5% for subject Q at 1640 knots. This is a large difference. However, when the four runs for each observer are averaged, the range from highest to lowest was only 64.82% to 87.96%. Thus, individual differences in percentage of false positives, averaged over trials, was not as large as differences found for some of the other performance indices examined in the present paper.

It is of some interest to see if there is any correlation between the number of targets detected by individuals and the number of objects that they mistake for targets. Do the numbers tend to vary together? When the data for all four speeds are combined and the Pearson product moment correlation coefficient, r, is computed it turns out to be + .673. This is statistically significant at the .01 level of probability. While not a large value of r, this indicates that numbers of detections and numbers of false positives do tend to vary together. Individuals who find many targets tend to find many false positives, those who find few of one tend to find few of the other, and intermediate scorers on the one tend to be intermediate on the other.

TABLE 13

PERCENTAGE OF FALSE POSITIVES

Observer	Aircraft Speed in Knots 700	1170	1640	2110	Overall Average +	Rank
A	78.49	78.57	78.49	81.90	79.36	15
B	73.44	78.26	82.35	84.57	79.66	16
C	89.66**	90.00**	85.71**	86.49**	87.96**	20
D	85.56	82.35	81.48	80.19	82.40	17
E	85.96	87.36	78.21	82.02	83.39	19
F	68.52	71.70	59.09	62.07*	65.34	2
G	65.00	68.97	64.10	67.50	66.39	5
H	73.21	56.52*	63.83	68.97	65.63	3
I	79.17	75.00	81.72	74.42	77.58	14
J	65.57	62.04	71.26	65.57	66.11	4
K	75.61	74.24	62.30	71.74	70.97	7
L	77.94	71.67	65.96	80.17	73.94	9
M	77.86	73.68	76.62	71.21	74.84	13
N	75.00	68.85	69.44	72.73	71.50	8
O	79.58	77.38	88.16	85.90	82.76	18
P	73.24	76.19	77.78	71.15	74.59	11
Q	68.63	76.36	57.50*	63.33	66.46	6
R	72.09	75.36	79.66	72.00	74.78	12
S	64.71*	61.70	70.37	62.50	64.82*	1
T	74.55	64.71	77.03	81.63	74.48	10
Mean	75.19	73.55	73.55	74.30	74.15	10
S.D.	6.91	8.31	9.06	8.02	6.99	
Ratio ++	1.39:1	1.59:1	1.49:1	1.39:1	1.36:1	

Prorated data was used as explained in the text.
+Average of the four percentages of false positive scores, i.e., row mean.
++Ration = ratio of highest score in the column to the lowest score.
*,** Lowest (best) score and highest (worst) score in the column, respectively.

The percentage of false positives for the four aircraft speeds, as indicated in the column means of table 13, do not appear to exhibit a trend with simulated aircraft speed. In other words, accuracy does not appear to be related to speed. To check this observation, all scores were arc sine transformed (since they are percentages) and Bartlett's test of homogeneity of variance was applied to the transformed scores. The chi-square, corrected for continuity, was only .653, far from the 9.5 required for statistical significance at the .05 level. Thus, the hypothesis of homogeneity of error variance (error variances essentially equal) is acceptable and analysis of variance is permissible. The analysis of interactions for percentage of false positives, given in table 14, reveals that neither Latin squares by trials interaction or Latin squares by velocities interaction was statistically significant, hence the Latin squares by velocities was lumped with the error sum of squares. In the Analysis of Variance (table 15) the velocity effect was not statistically significant, verifying the earlier observation that different aircraft speeds do not result in greater differences in accuracy or in its linear transform, percentage of false positives, than would be expected on the basis of chance. A nonparametric test of the hypothesis of no velocity effect on percentage of false positives, the Friedman Two-Way Analysis of Variance by Ranks, gave an χ^2_r with three degrees of freedom of only 1.995, which was not statistically significant. Thus, the two analyses agree.

It is also of interest to note that the differences between subjects in the same Latin square are statistically significant at the .001 level. Trials and Latin squares were also significantly different. It must be concluded that some sort of learning (or fatigue or whatever) effect is present.

TABLE 14

ANALYSIS OF INTERACTIONS FOR PERCENTAGE+ OF FALSE POSITIVES

Source of Variation	Sum of Squares	df	Mean Square	*F*
Latin Squares x Trials	111.82	12	9.318	1.30
Latin Squares x Velocities	86.51	12	7.209	.95
Error	227.26	30	7.575	

+ Arc sin transformed data was used.
NOTE: Neither interaction is statistically significant at the .05 level.

TABLE 15

ANALYSIS OF VARIANCE OF PERCENTAGE OF FALSE POSITIVES+

Source of Variation	Sum of Squares	df	Mean Square	*F*
Latin Squares	777.19	4	194.30	26.01***
Between Subjects in Same Square	896.12	15	59.74	8.00***
Velocities	11.69	3	3.897	.52
Trials	119.47	3	39.82	5.33**
Latin Squares x Trials	111.82	12	9.318	1.25
Error (including LS x Velocities)	313.77	42	7.471	
TOTAL		79		

+Arc sin transformed data was used.
***Statistically significant at the .001 level.
**Statistically significant at the .01 level.

D. MEANING AND MEASUREMENT OF SCREEN POSITION OF DETECTED OBJECTS

A radar observer looking for targets on a display will not detect a target the instant that it appears on the display: he notices it only after it has been on the display for a time. In the moving-image display on an SLR system, this lag in detection means that the image of a target has moved down some distance from the top edge of the display before it is detected. Target images move down the screen at a rate proportional to the ground speed of the aircraft. Thus, screen position of a target when it is detected, i.e., screen travel, is linearly related to the time lag between the appearance of a target on the display and its detection by an observer, and to the distance along the ground that the aircraft has traveled in this time interval. Distance relationships for side-looking radar are shown in figure 1.

Since quick detection may be necessary for launching a successful attack, screen position on the display of detected targets must be utilized in assessing the performance of both the observer and the system that includes him.

The screen position (or image travel) of objects identified by subjects as targets was obtained by projecting the pictures obtained by the data camera onto a screen. The screen was marked off into 11 intervals representing equal distances on the original SLR display. Thus, the data represent elevenths of the 14-inch (356 mm) screen height. Since the screen height corresponds to 41.5 nautical miles of terrain, elevenths of screen height may be converted to nautical miles by multiplying by 41.5/11, i.e., by 3.77.

E. SCREEN POSITION OF DETECTED TARGETS

Table 16 gives the average screen travel of detected targets for individual subjects in elevenths of screen height. Examination of the column of means on the right of the table reveals a wide range of scores. Thus, subject "E" has an average overall for simulated aircraft speeds of 2.822 while subject "N" has 7.486. This range of 2.7 to 1 corresponds to an average of 10.6 nautical miles versus 28.2 nautical miles of aircraft travel between display and detection of targets. At every speed test subject "N" requires two to three times as much aircraft travel as subject "E". Note that the range ratios at the bottom of the velocity columns range from 2.58:1 to 4.34:1. Other comparisons may be more easily visualized and the variability of the data may be made more obvious by examination of the graph shown in figure 14 than from studying the table. The graph clearly shows that some individuals are much slower, on the average, than others. Also, most individuals vary greatly from one speed to another and from one trial to another. It will be shown later on in this paper that only a very small portion of the large variation from trial to trial is due to a difference in simulated aircraft speed.

The large variation in performance from trial to trial means that it will be difficult, if not impossible, to devise, for crew selection or training purposes, only one or two short tests that can reliably and accurately rank observers for the rapidity with which they can be expected to find targets. Extensive testing may be necessary.

TABLE 16

AVERAGE DISTANCE+ TARGETS TRAVEL DOWN THE DISPLAY
BEFORE BEING DETECTED

Subjects	700	1170	1640	2110	Mean	Rank
	Aircraft Speed in Knots					
A	2.650	3.762	5.700	5.048	4.290	14
B	3.412	3.267	3.200	4.720	3.650	7
C	3.917	7.000	5.800	8.600**	6.329	18
D	4.185	4.667	7.167	6.333	5.588	17
E	2.167	3.000	3.059	3.063	2.822*	1
F	4.588	4.267	5.000	5.364	4.805	15
G	4.357	3.667	4.143	4.538	4.176	13
H	4.200	3.900	3.176	3.611	3.722	8
I	2.533	4.700	3.824	3.909	3.742	9
J	3.143	4.474	2.920*	3.048*	3.396	6
K	2.800	1.824*	3.348	4.154	3.032	2
L	3.000	4.824	7.063	4.957	4.961	16
M	2.759	2.100	3.111	5.474	3.361	5
N	7.250**	7.316	7.545**	7.833	7.486**	20
O	1.923*	4.000	3.556	3.273	3.188	3
P	2.316	3.200	3.750	3.533	3.200	4
Q	2.875	4.231	4.353	3.909	3.842	10
R	3.167	4.353	3.833	5.071	4.106	12
S	3.667	3.944	3.688	4.333	3.908	11
T	6.571	7.917**	6.353	4.222	6.266	19
Mean	3.574	4.321	4.529	4.750	4.294	
S.D.	1.371	1.557	1.525	1.468	1.255	
Range Ratio ++	3.77:1	4.34:1	2.58:1	2.82:1	2.65:1	

+The distances in this table are in elevenths of the screen height. To convert to nautical miles, multiply by 3.77.
++Range ratio is the ratio of the maximum to the minimum score
*Smallest score in column
**Largest score in column

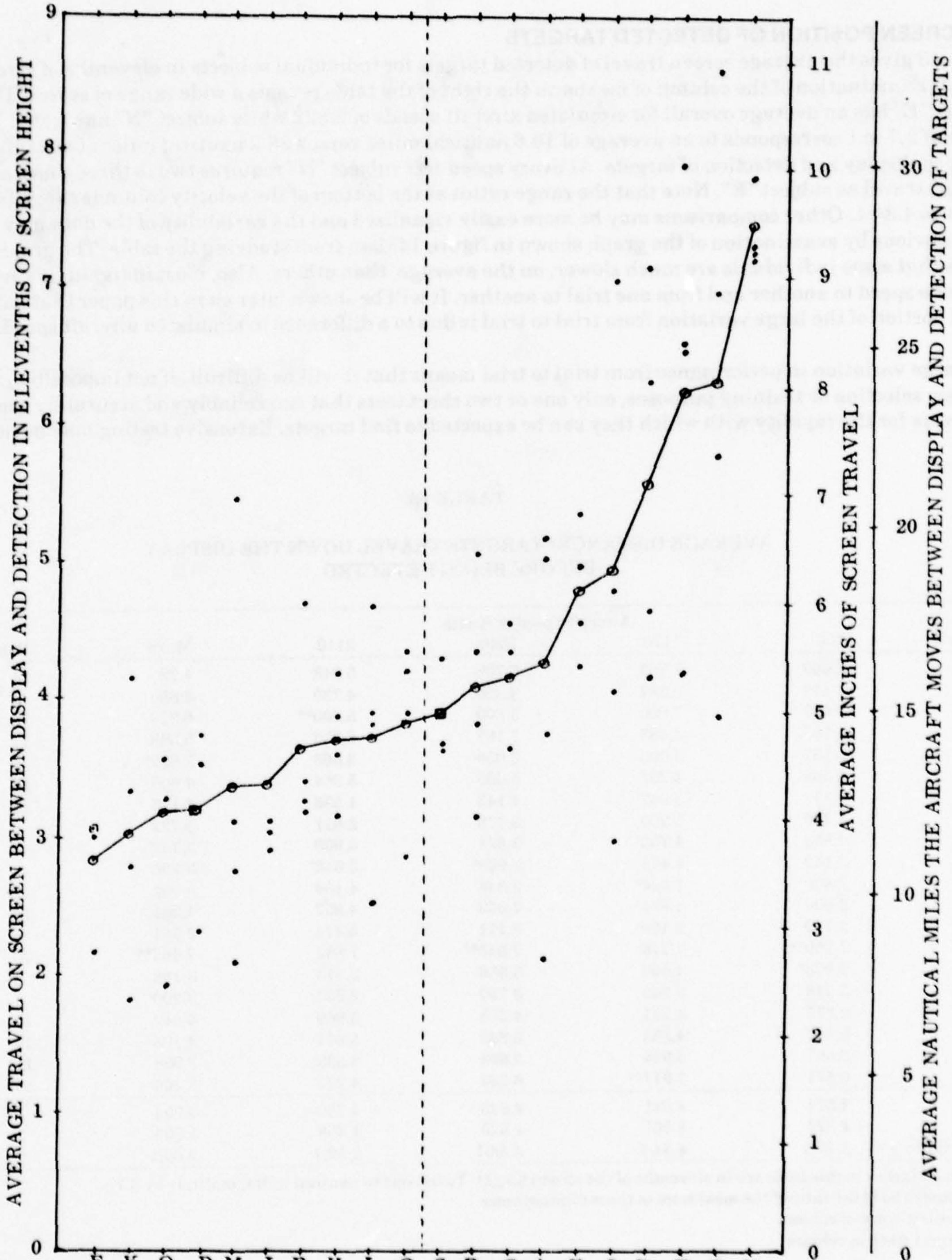


Figure 14. Individual Differences in Distance Traveled Between the Appearance on the Display of a Target and Its Detection by the Observer.

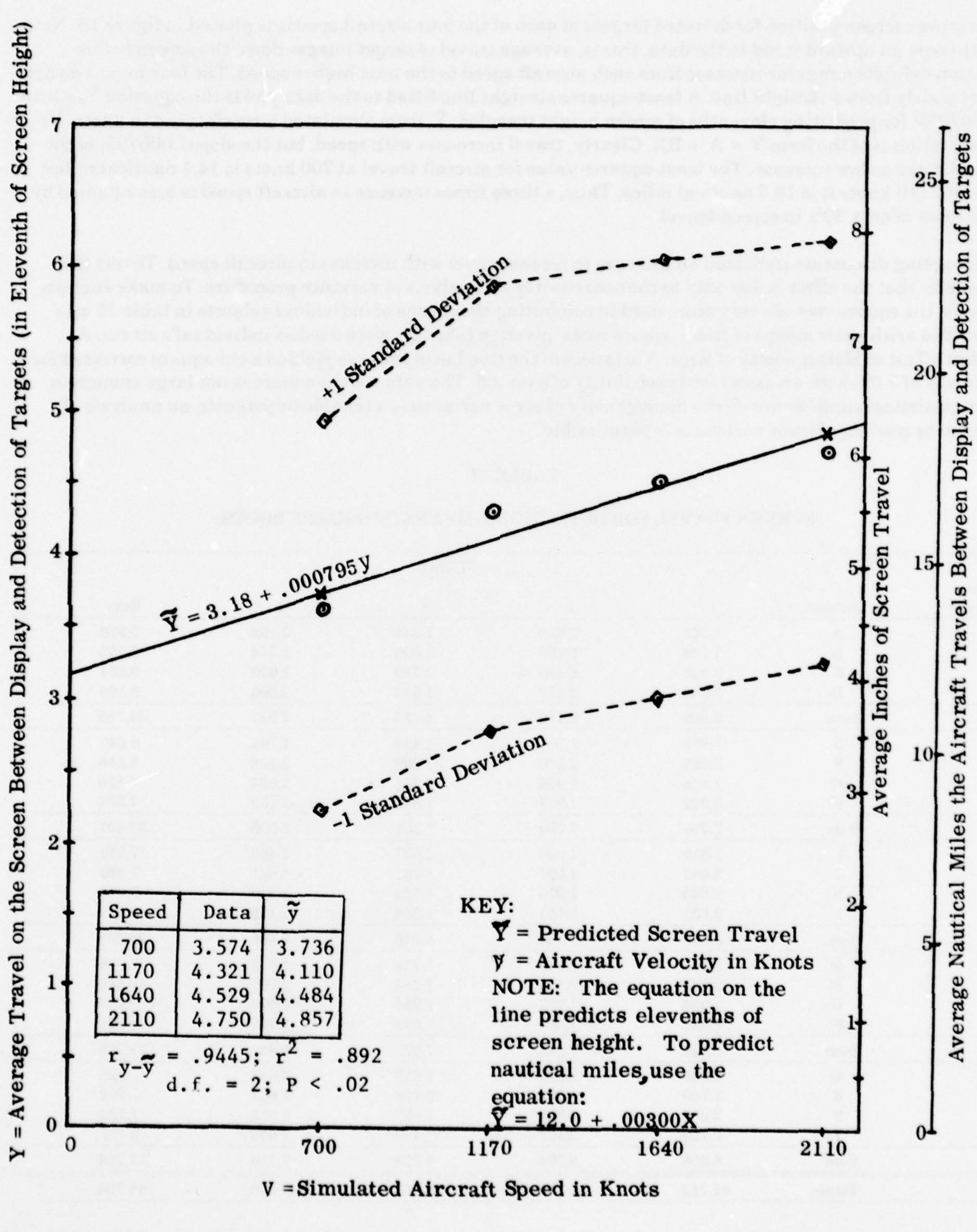


Figure 15. The interval between display and detection of targets at the four different simulated aircraft speeds.

The average screen position for detected targets at each of the four aircraft speeds is plotted in figure 15. Note that there is an upward trend to the data, this is, average travel of target images down the screen before detection exhibits a regular increase from each aircraft speed to the next higher speed. The four means do not depart widely from a straight line. A least-squares straight line fitted to the data yields the equation $\hat{Y} = 3.18 + .000795V$ for predicting elevenths of screen height traveled, $\hat{Y}$, from simulated aircraft speed in knots, V. This equation is of the form $Y = A + BX$. Clearly, travel increases with speed, but the slope (.000795) of the line indicates a slow increase. The least-squares value for aircraft travel at 700 knots is 14.1 nautical miles, while at 2110 knots it is 18.3 nautical miles. Thus, a three times increase in aircraft speed is accompanied by an increase of only 30% in screen travel.

The foregoing discussion indicated an increase in screen travel with increase in aircraft speed. To test the hypothesis that the effect is due only to chance requires an analysis of variance procedure. To make such an analysis, the square root of every score used in computing the means of individual subjects in table 16 was taken. The arithmetic means of these square roots, given in table 17, were used as individual's scores. A Bartlett's Test of Homogeneity of Error Variances of the five Latin squares yielded a chi-square corrected for continuity of 7.07, with an associated probability of over .05. The value of chi-square is not large enough to attain statistical significance. Since homogeneity of error variance is a tenable hypothesis, an analysis of variance by pooling of error variances is permissible.

TABLE 17

SCREEN TRAVEL FOR DETECTIONS (MEANS OF SQUARE ROOTS)

Latin Square	Subject	TRIAL 1	2	3	4	Sum
1	A	1.565	2.328	1.834	2.193	7.920
	B	1.749	1.766	2.096	1.744	7.355
	C	2.925	2.610	2.389	1.910	9.834
	D	2.660	2.476	1.934	2.090	9.160
	Sum	8.899	9.180	8.253	7.937	34.269
2	E	1.693	1.704	1.439	1.704	6.540
	F	2.269	2.170	1.969	2.108	8.516
	G	1.815	1.998	1.953	2.054	7.820
	H	2.022	1.927	1.843	1.739	7.531
	Sum	7.799	7.799	7.204	7.605	30.407
3	I	1.912	1.920	2.097	1.460	7.389
	J	2.057	1.697	1.701	1.641	7.096
	K	1.650	1.970	1.784	1.310	6.714
	L	2.631	2.151	1.366	2.185	8.333
	Sum	8.250	7.738	6.948	6.596	29.532
4	M	2.307	1.579	1.734	1.418	7.038
	N	2.674	2.734	2.781	2.667	10.856
	O	1.364	1.737	1.954	1.872	6.927
	P	1.883	1.701	1.453	1.855	6.892
	Sum	8.228	7.751	7.922	7.812	31.713
5	Q	2.040	1.948	1.619	2.018	7.625
	R	1.700	1.961	2.178	1.863	7.702
	S	2.039	1.866	1.952	1.882	7.739
	T	2.759	2.523	2.479	1.956	9.717
	Sum	8.538	8.298	8.228	7.719	32.783
1-5	Totals	41.714	40.766	38.555	37.669	158.704

Table 18 gives the results of the analysis of variance for screen travel, i.e., screen position when detected, of the interactions using square root transformed data. Neither the Latin squares-by-trials interaction nor the Latin squares-by-velocities interaction is statistically significant.

By pooling the sums of squares and degrees of freedom for error and for the (Latin squares) χ (velocities) interaction, the best common estimate of error is obtained. Table 19 gives the analysis of variance using this error term. Note that screen positions at detection for the different velocities are significantly different at the .001 level of significance. In fact, the probability of obtaining by chance alone such a large "F", 65.3, as that found in the analysis is considerably less than one in a thousand. Thus, the earlier finding that screen travel increases with aircraft speed is verified. The F significant at the .01 level for differences between subjects in the same Latin square for screen position is consistent with the large individual differences that were clearly apparent from inspection of figure 14.

The significant F for Latin squares is probably largely due to sampling: large differences between individuals and large differences within the same individuals from one test session to the next. These differences are large compared to differences attributable to the different simulated aircraft speeds.

The analysis of variance and the screen position plot just discussed have examined only the means of the distributions of number of detections. The percentage of available targets detected in each of eleven equal-sized intervals down the display screen gives some insight as to the nature of the distribution of target detections on the display. The data are given in table 20 for each of the four aircraft speeds. They are plotted in figure 16. Note that the number and percentage of targets, which is the vertical dimension on the plot, is logarithmic. Note that, in the first eleventh of the screen height, the number of targets detected falls rapidly with an increase in aircraft speed. Inspection and decision times are likely responsible for the poorer

TABLE 18

SCREEN TRAVEL+ FOR DETECTED TARGETS: ANALYSIS OF INTERACTIONS

Source of Variation	Sum of Squares	df	Mean Square	F
Latin Squares x Trials	.317	12	.0264	.426
Latin Squares x Velocities	.709	12	.0591	.953
Error	1.859	30	.0620	

NOTE: Neither interaction is statistically significant.
+ Square root transformed data was used.

TABLE 19

SCREEN TRAVEL+ FOR DETECTED TARGETS: ANALYSIS OF VARIANCE

Source of Variation	Sum of Squares	df	Mean Square	F
Latin Squares	.884	4	.221	3.56*
Between Subjects in Same Square	5.457	15	.364	5.86**
Velocities	1.196	3	.399	65.3**
Trials	.531	3	.177	2.85*
Squares x Trials	.317	12	.0264	.43
Error (Including LS x Velocities)	2.568	42	.0621	
TOTAL	10.953	79		

*Significant at the .05 level.
**Significant at the .01 level.
+Square root transformed data was used.

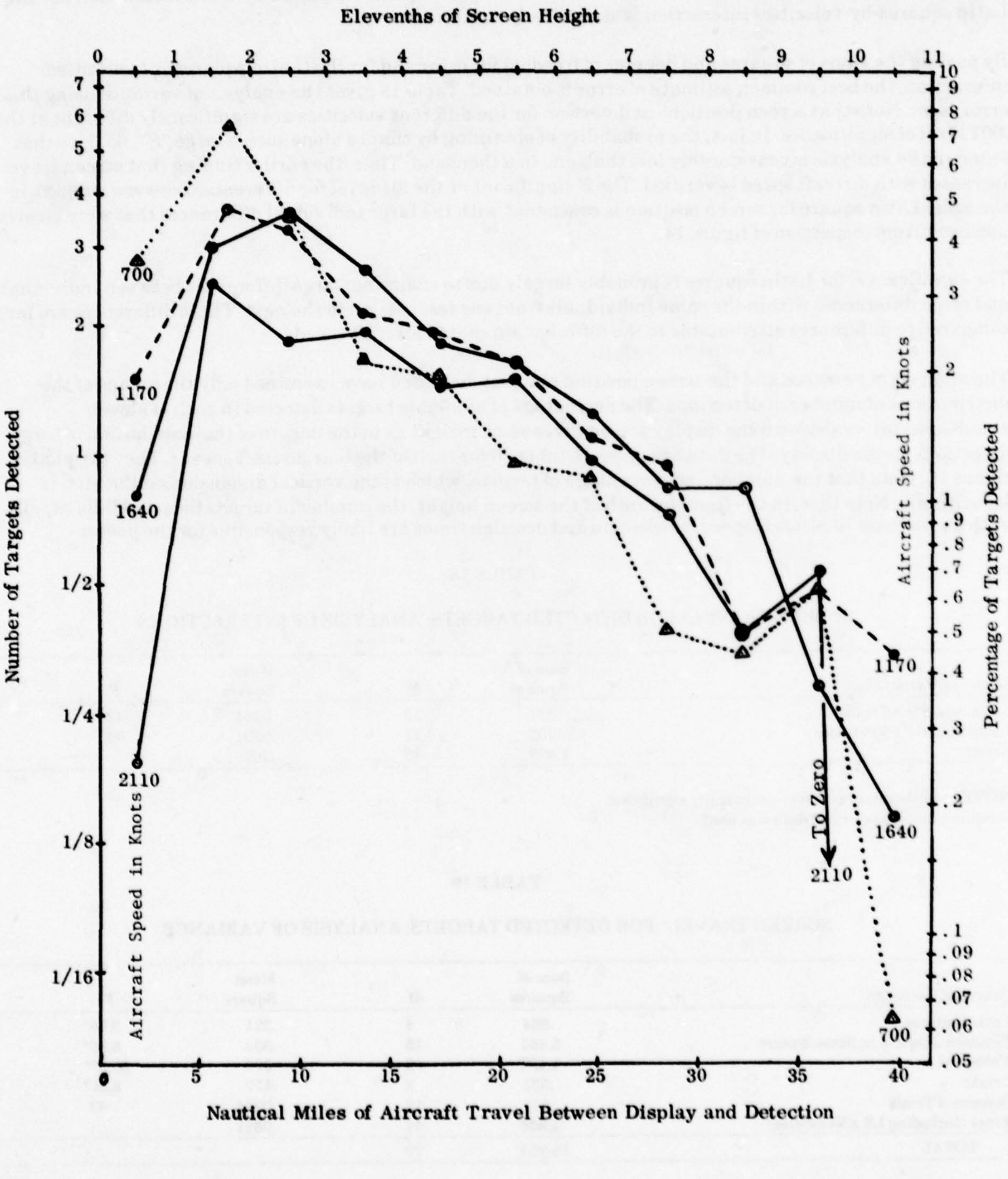


Figure 16. Number of targets detected as a function of ground distance covered between the display and detection of targets.

performance at the higher speeds in the first interval. In the first two intervals an advantage in number of detections is gained at the lower speeds. This point is made very plain by summing the data to obtain a cumulative frequency distribution for each aircraft speed. The four distributions are plotted in figure 17. Any point on any one of the four curves indicates how many targets had been detected up to that position on the screen. The last point on the right side of each curve indicates the total number of targets detected at the speed represented by the curve.

The curves are cleanly separated and the number of targets detected by any given fraction of the way down the display screen is lower at any given simulated aircraft speed than at any slower speed. The already-noted advantage gained in the first interval or so at a slower speed is maintained or even increased. On all four curves there is a rapid increase up to about half of the screen height, then all curves are characterized by a decreasing rate of increase, although none level off entirely. In figure 18, cumulative percentage of detections is plotted against time rather than distance on the display or on the ground. Note that the curves for the two highest speeds are almost touching over most of their lengths, i.e., for the first approximately 60 seconds the total number of detections as a function of time is almost the same at the two fastest speeds. Note also that the four curves lie in order from highest to lowest in going from left to right, i.e., along the "time" axis.

TABLE 20

PERCENTAGE OF AVAILABLE TARGETS DETECTED IN EACH OF ELEVEN EQUAL INTERVALS DOWN THE DISPLAY SCREEN

		Percentage of Available Targets Detected			
Display Interval	Interval Center	Aircraft Speed in Knots 700	1170	1640	2110
0-1	.5	3.59	1.92	1.03	.32
1-2	1.5	7.37	4.81	3.91	4.17
2-3	2.5	4.62	4.29	4.74	3.01
3-4	3.5	2.12	2.76	3.46	3.21
4-5	4.5	1.99	2.50	2.31	2.37
5-6	5.5	1.22	2.12	2.12	2.50
6-7	6.5	1.15	1.41	1.60	1.60
7-8	7.5	.51	1.22	1.09	1.22
8-9	8.5	.45	.51	1.09	.64
9-10	9.5	.64	.64	.38	.90
10-11	10.5	.064	.45	.19	0

NOTE: The data in this table were not prorated, hence the percentages are slightly lower than those in prorated tables.

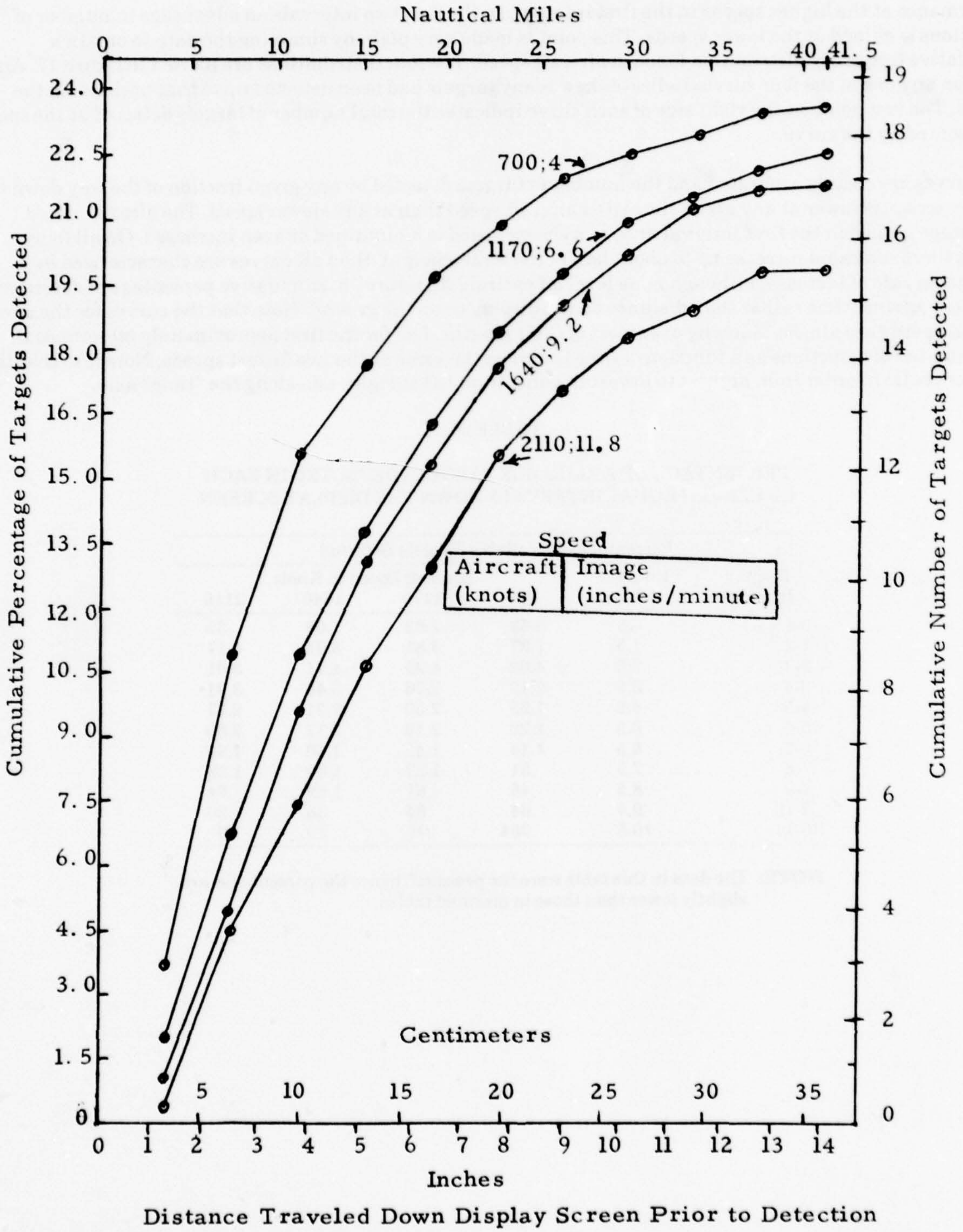


Figure 17. Cumulative Radar Target Detections on the Display Screen at the Four Simulated Aircraft Speeds.

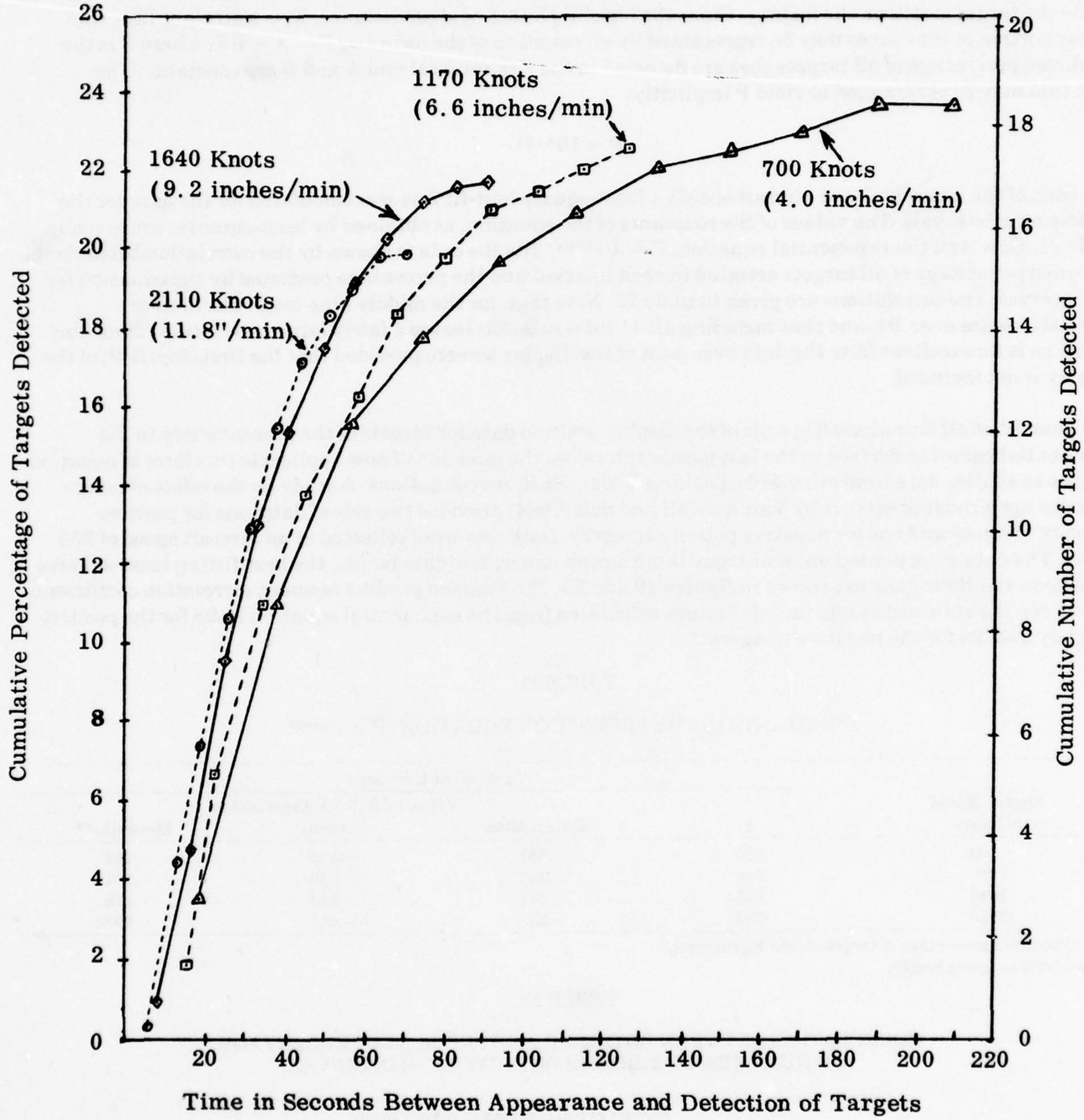


Figure 18. Cumulative Percentage of Targets Detected on a 14 x 14 Inch Display of Side-Looking Radar Imagery at a Scale of 1:216,000 as Functions of the Time Interval Between Initial Appearance on the Display and Detection by the Subject for Four Different Simulated Aircraft Speeds

Going back to figure 16, it is apparent that, from the second through the tenth interval, the curves are almost linear. The very small number of targets detected in the eleventh interval is due in part to the inability of the subject, when a target is spotted in this interval, to press all appropriate scoring buttons and record a detection while the target is still on the display. Since the log of P plotted against distance, X, is a straight line, the linear portion of the curves may be represented by an equation of the form Log P = A + BX, where P is the predicted percentage of all targets that are detected in the Xth interval and A and B are constants. The equation may be rearranged to yield P implicitly:

$$P = 10^{A+BX}$$

For each of the four simulated aircraft speeds a least-square best-fit line was calculated for the data for the middle nine intervals. The values of the constants of the equation, as obtained by least-squares, are given in table 21. How well the exponential equation, $P = 10^{A+BX}$, fits the data is shown by the correlation between the obtained percentage of all targets detected in each interval and the percentage predicted by the equation for the interval. The correlations are given in table 22. Note that, for the middle nine intervals, all four correlations are over .94, and that including all 11 intervals still leaves a fairly large correlation. Thus, the equation is an excellent fit to the data over most of the display screen, provided that the first (top) fifth of the display is not included.

The close fit at all four aircraft speeds of the display position data for targets of the present study to the exponential equation derived in the last paragraph raises the question of how applicable this form of equation may be to similar data from other Side-Looking Radar (SLR) investigations. A study on the effect of image polarity upon radar observers by Van Ausdall and Self (1964) provides two sets of data, one for positive polarity imagery and one for negative polarity imagery. Both sets were collected at an aircraft speed of 950 knots. The data were plotted on semi-logarithmic graph paper. The data points, the best-fitting least-squares equations and their plots are shown in figures 19 and 20. The Pearson product moment correlation coefficient, r, between the obtained values and the values calculated from the exponential equation is .98 for the positive imagery and .96 for the negative imagery.

TABLE 21

CONSTANTS IN THE PREDICTION EQUATION* $\hat{P} = 10^{A+BX}$

Aircraft Speed in Knots	Value of the Constants: A	Value of "B" for X Expressed in: Nautical Miles	Seconds	Elevenths**
700	.966	−.551	−2.83	−.146
1170	.910	−.451	−1.40	−.121
1640	.915	−.448	−.984	−.119
2110	.792	−.358	−.611	−.0949

*P = Predicted percentage of targets in the X_{th} interval.
**Elevenths of screen height.

TABLE 22

CORRELATION BETWEEN OBTAINED AND PREDICTED PERCENTAGE (OR NUMBERS) OF TARGETS DETECTED PER INTERVAL

Aircraft Speed	Number of Intervals out of Eleven Used: All Eleven	Middle Nine
700	.916	.954
1170	.885	.971
1640	.761	.948
2110	*	.947
950		.981** + imagery
950		.960** − imagery

*The last interval had a P of 0, thus it was far off from a straight line.
**From a previous study by Van Ausdall and Self, 1964, see figures 19 and 20.

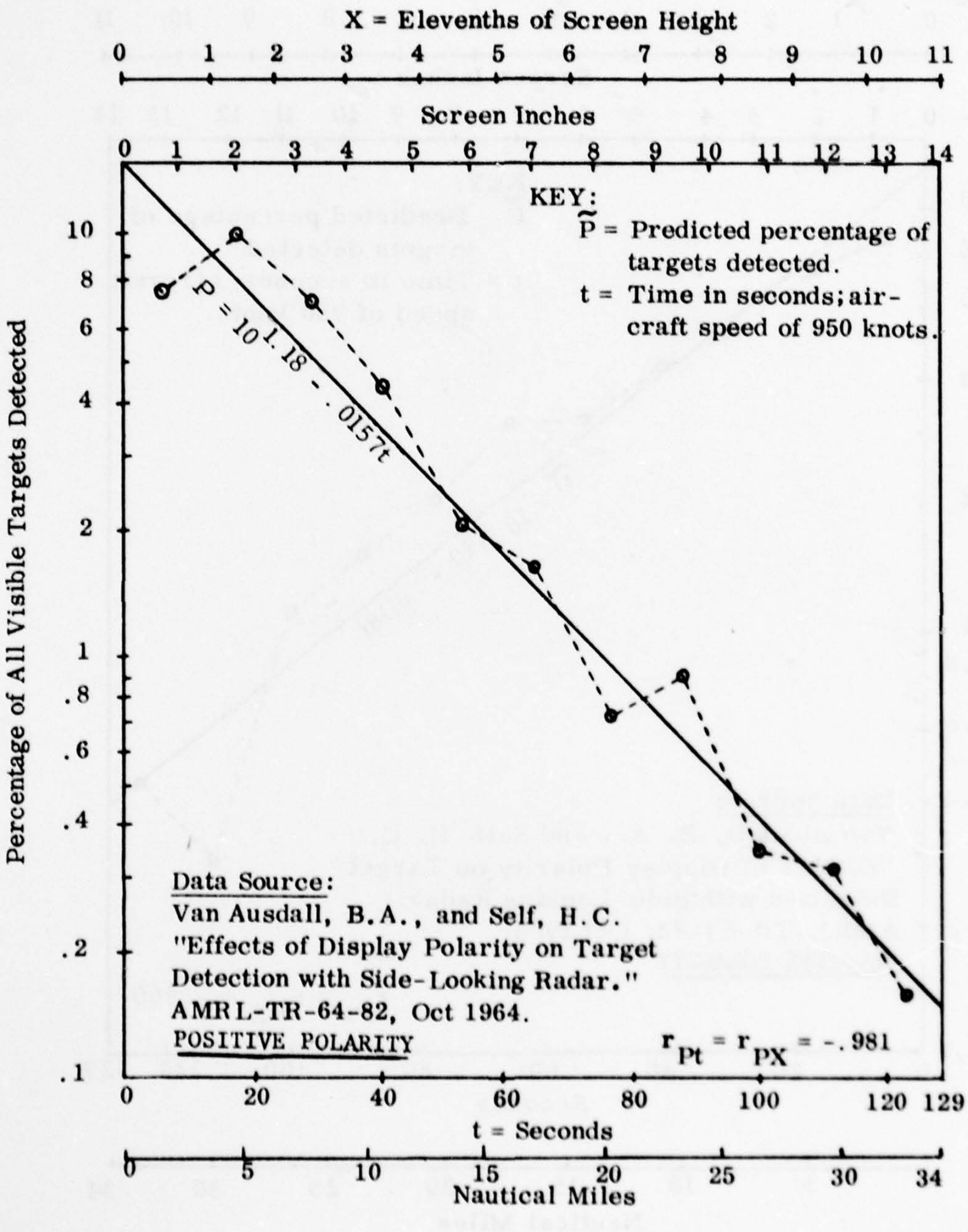


Figure 19. Percentage of Targets Detected as a Function of the Interval Between Display and Detection. Data for Positive Image Group in Reference.

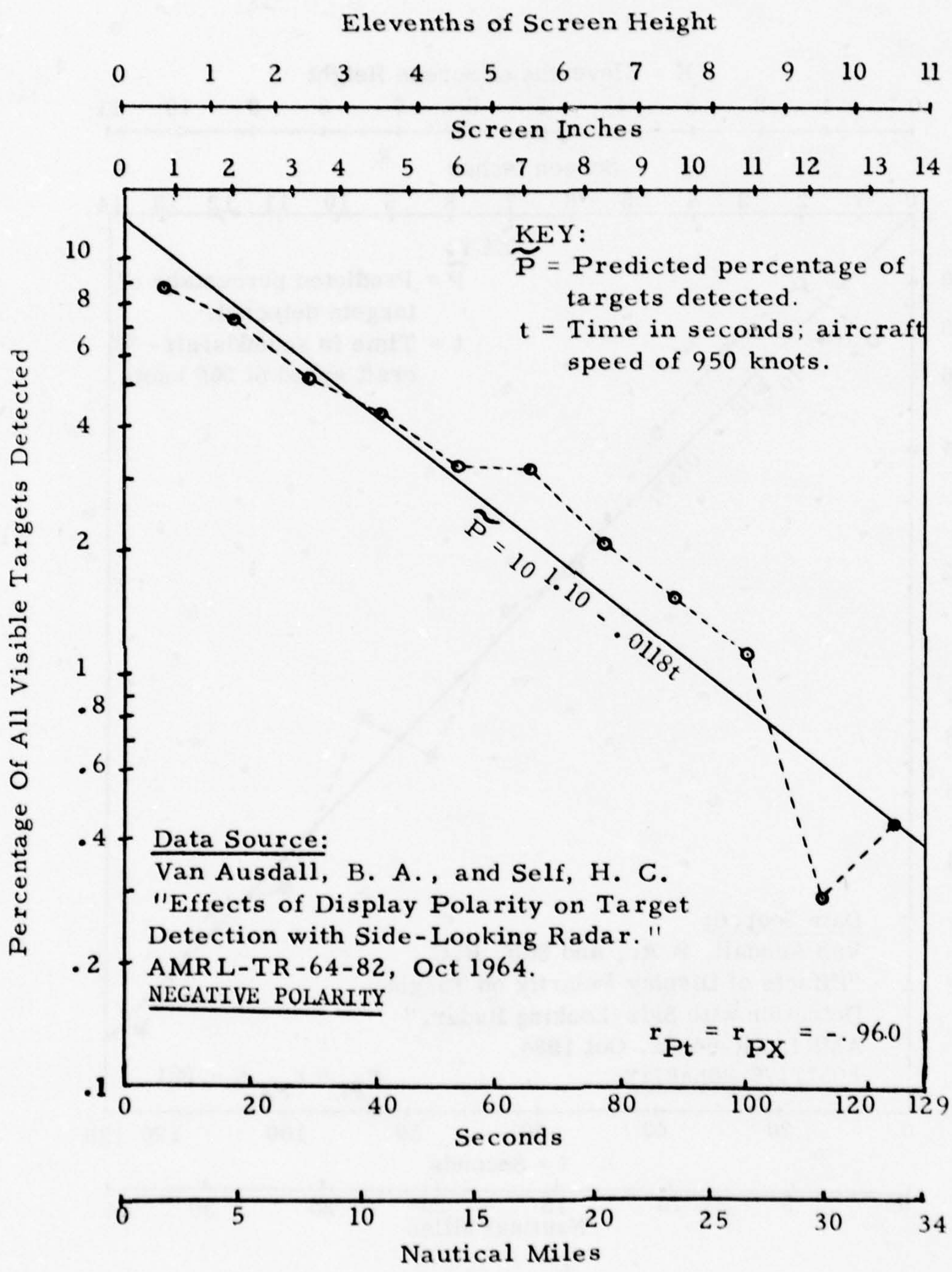


Figure 20. Percentage of targets detected as a function of the interval between display and detection. Data for negative image group in reference.

An additional set of data for SLR, provided by Self and Bate (1969), is plotted in figure 21. The aircraft speed in this study was 1320 knots. The value of r for this data is .9896 for the bottom or last 9 of the 11 equal intervals down the display screen.

It is clear that the exponential equation has some generality: it is applicable to data from various studies. The number or percentage of targets detected is related to the position on the display when detected (or to time to detect or to aircraft travel over the terrain after target images appear on the display) by an equation of the form $P = 10^{A + BX}$, except at the top of the display.

The values of the constant "B" in table 22 decrease as aircraft velocity increases. The values of the constant are plotted in figure 22 where a least-squares best-fitting straight line, $B = .628 - .000125V$, has been computed for the data, with V being aircraft velocity in knots. Here, B is the slope of the equation $\text{Log } P = A + BX$ plotted on semi-log paper, when X is elevenths of display screen height. As may be seen from inspection of table 22 and figure 22, due to the large drop in the value of the constant "A" at 2110 knots, it does not appear that this constant is linearly related to aircraft speed, V.

The above results on display position at detection (or screen travel) may be summarized by noting that the percentage of targets detected, P, is related to position, X, by an exponential equation of the form $P = 10^{A + BX}$ (or $P = e^{C + DX}$). Further, the constant "B" is, in the present study, linear with V, aircraft speed, so that $P = 10^{A+CX-DVX}$. The variable "X" may be either a display screen position measure, or a measure of terrain distance overflown, and, with the appropriate constants, indicates either travel of the image or of the aircraft between the initial display of the target and its designation (detection) by an observer. It must be kept in mind that the above results, in particular the equations, from the various studies, were derived from Side-Looking Radar images. SLR is a "mapping" sensor in that the image scale is the same everywhere on the display, a condition not found with some sensors, such as TV, when not aimed straight down.

F. SCREEN POSITION FOR FALSE POSITIVES

The average screen position (or screen travel prior to response) for false positives for individual observers is given in table 23. Examination of this table and the graph of the data in figure 23 reveals that, as was the case with real targets, the differences between individual observers are, on the average, quite large. For every simulated aircraft speed, the ratio of the average screen travel for the slowest reacting observer to that of the most rapid responder was 2.5:1 or greater. Note from the table that at 1170 knots the ratio was 4.3:1. Also, from the graph it is apparent that the large differences are not merely due to the presence at the ends of the range of one or two observers who are exceptional in speed. Also striking is the fairly large variation from trial-to-trial for most individuals.

The variability in screen position at detection or amount of travel down the screen prior to detection appears to be about the same as was the case for detected targets. The *F*, or variance ratios, for the various simulated aircraft speeds, going from slowest to fastest and with target variances in the numerator and false positive variances in the denominator are, respectively, 1.02, 1.11, 1.17, and 1.24. With 18 degrees of freedom in both numerator and denominator, it is clear that none of these *F* ratios even approach statistical significance. It is concluded that the hypothesis of equality of variances in target travel for targets and for false positives is acceptable.

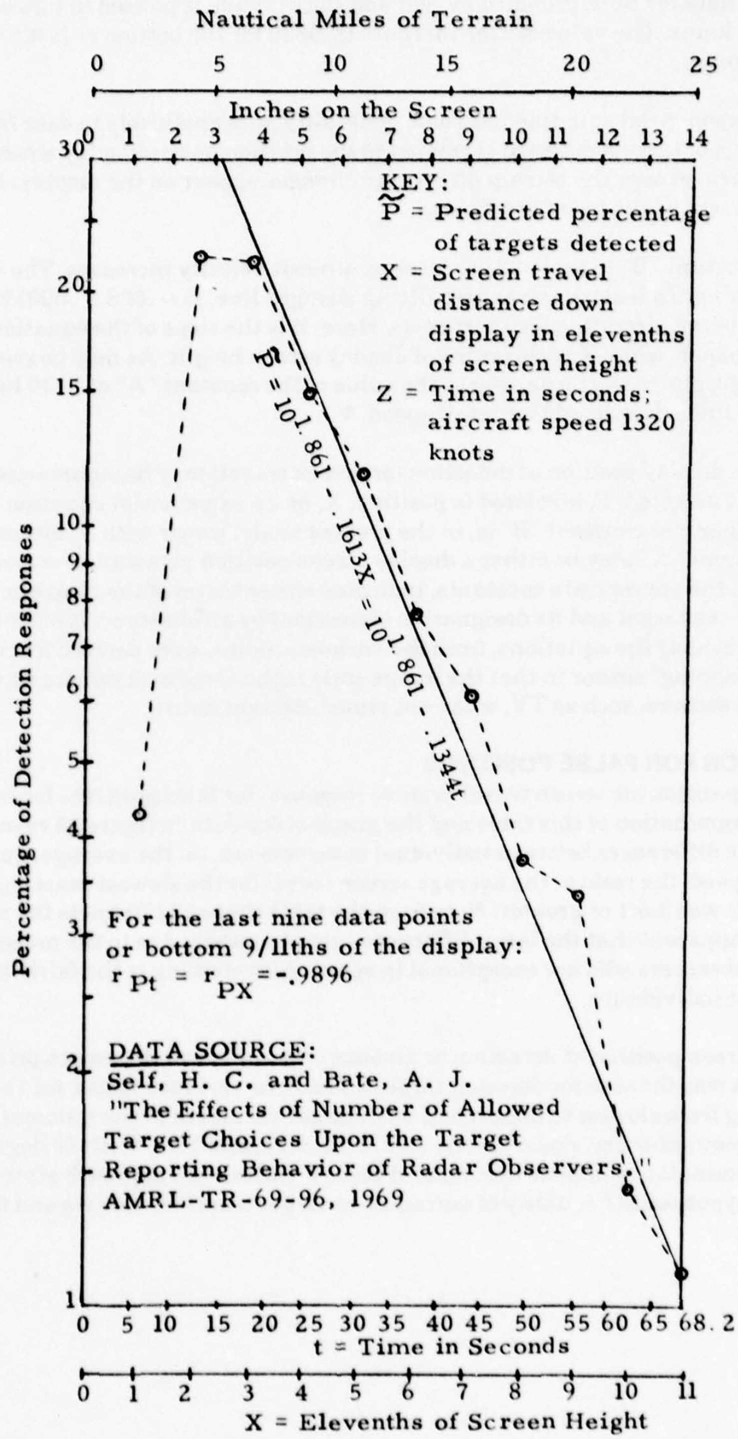
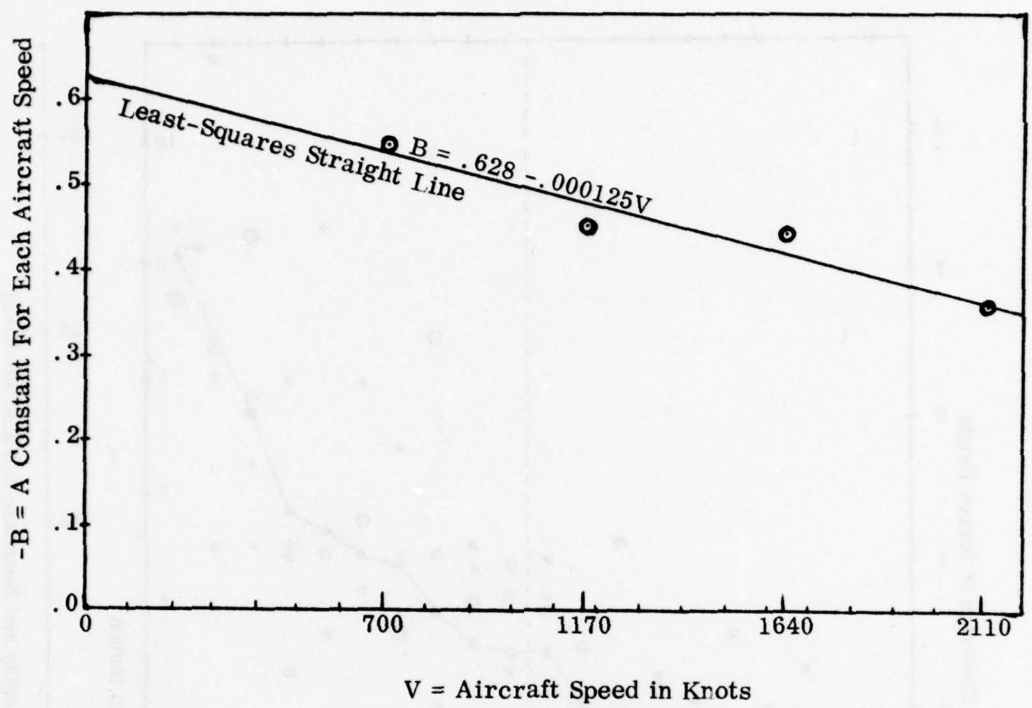


Figure 21. Percentage of Targets Detected as a Function of the Interval Between Display and Detection. Data from a Previous Study by Self and Bate, for all Conditions Combined.



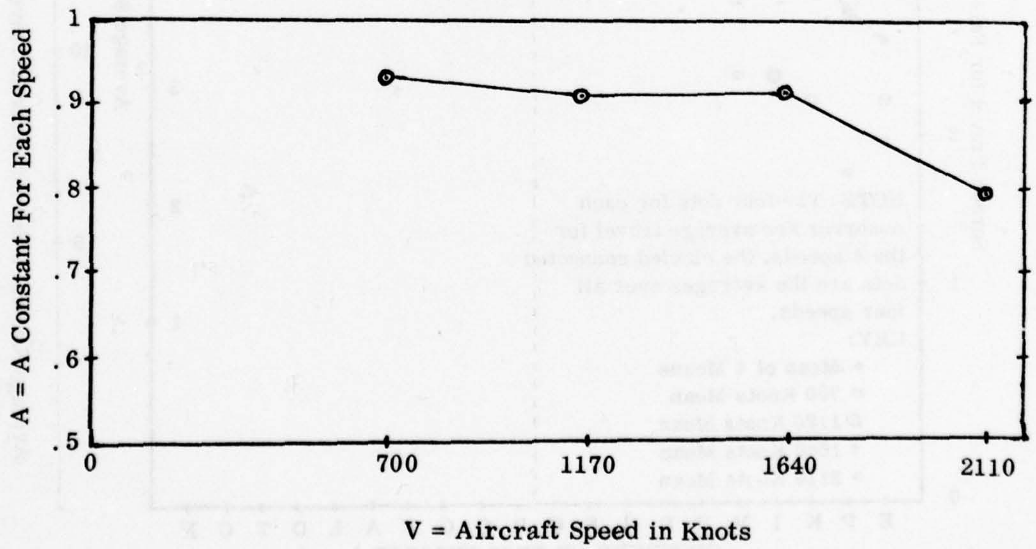


Figure 22. The Constants A and B as a Function of Aircraft Speed in the Exponential Equation $P = 10^{A+BX}$ Relating the Predicted Percentage of Targets Detected to Distance Down the Display Where Detection Occurred, i.e., Relating %D and Screen Travel.

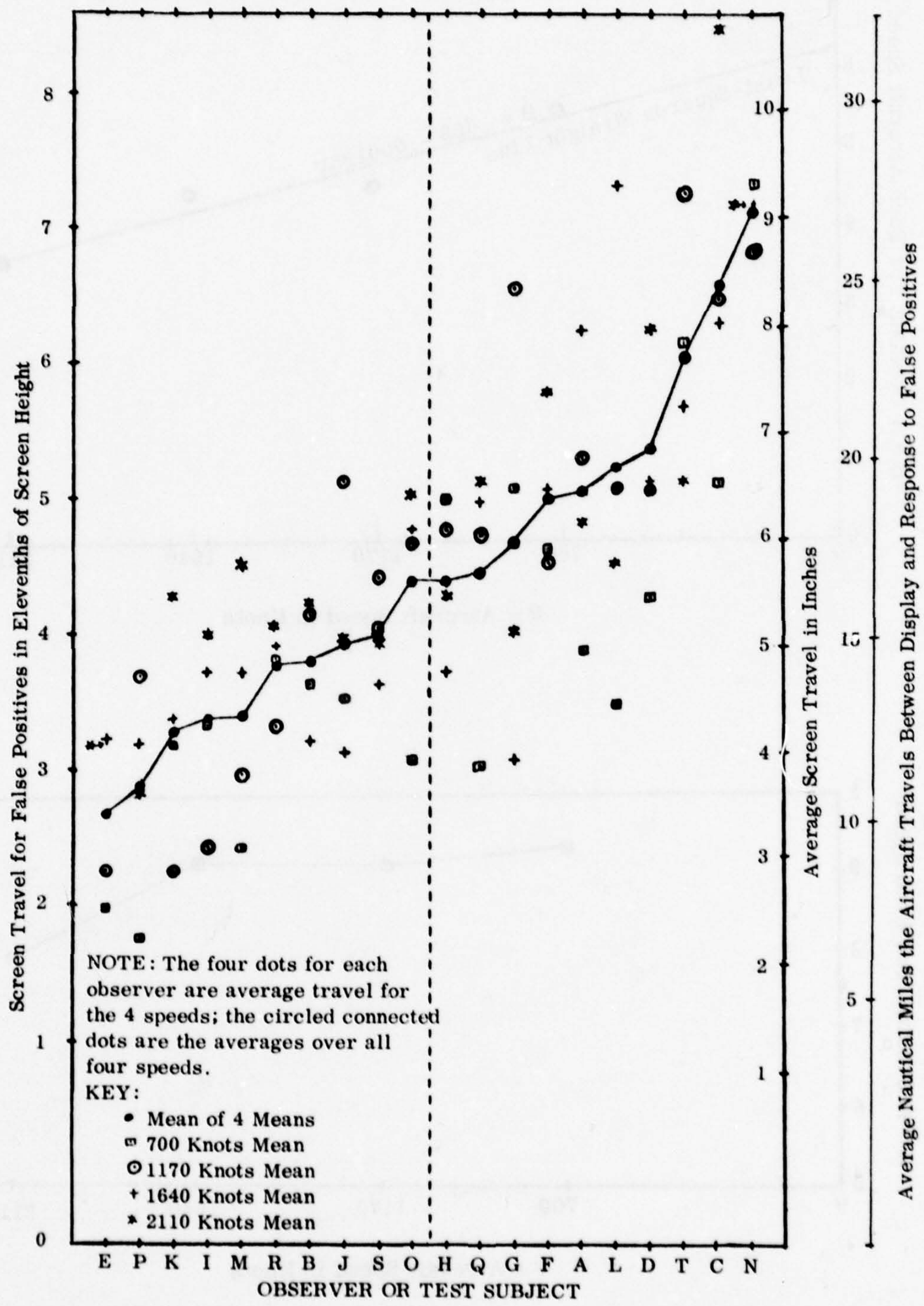


Figure 23. Individual differences in distance traveled between the appearance of a nontarget mistaken for a target and response to it by the observer.

TABLE 23

AVERAGE DISTANCE* NONTARGETS TRAVEL DOWN THE DISPLAY BEFORE BEING CALLED TARGETS

SUBJECT	Aircraft Speed in Knots 700	1170	1640	2110	MEAN	RANK
A	3.86	5.30	6.24	4.84	5.06	15
B	3.62	4.15	3.21	4.24	3.81	7
C	5.12	6.48	6.30	8.47	6.59	19
D	4.27	5.13	5.80	6.25	5.36	17
E	1.99	2.25	3.23	3.21	2.67	1
F	4.62	4.53	5.08	5.78	5.00	14
G	5.08	6.55	3.08	4.04	4.69	13
H	5.00	4.77	3.73	4.28	4.45	11
I	2.44	3.33	3.74	4.00	3.38	4
J	3.52	5.13	3.13	3.95	3.93	8
K	3.19	2.25	3.39	4.27	3.27	3
L	3.49	5.60	7.32	4.57	5.25	16
M	2.96	2.43	3.72	4.51	3.41	5
N	7.33	6.83	7.16	7.16	7.12	20
O	3.06	4.69	4.78	5.04	4.39	10
P	1.75	3.69	3.19	2.84	2.87	2
Q	3.03	4.74	4.96	5.11	4.46	12
R	3.81	3.33	3.91	4.06	3.78	6
S	4.03	4.41	3.61	3.97	4.01	9
T	6.15	7.25	5.68	5.13	6.05	18
Mean	3.92	4.64	4.56	4.79	4.48	
S.D.	1.36	1.48	1.41	1.32	1.19	
Range Ratio	3.78:1	4.34:1	2.45:1	2.82:1	2.67:1	

*Tabled values are in elevenths of screen height. To convert to inches on the screen, multiply by 1.27. If nautical miles on the terrain are desired, multiply tabled values by 3.77.

Bartlett's test for homogeneity of variance of the square-root-transformed screen travel data for false positives at different aircraft speeds yielded a chi-square corrected for continuity of 3.50, which did not approach significance at the .05 level. It is concluded that the four variances are not significantly different from each other, and that analysis of variance may thus be performed on the data. Table 24 gives the analysis of interactions. Since neither interaction is positive, the Latin squares by velocities sum of squares was pooled with the error sum of squares from the Latin squares to obtain the best common estimate of error which can be obtained from the data. The analysis of variance of table 25 yields a highly significant ($P<.001$) difference between subjects in the same Latin square, a result expected from the earlier discussion of differences between individual observers. It is concluded that the large observer differences are not all attributable to chance. Trials were also significantly different at the .001 level of statistical significance. The main interest of the present study is in the velocity effect, and it was highly significant ($P<.001$), leading to the conclusion that screen position at detection or screen travel prior to observer response for false positives varies significantly with aircraft velocity. A Friedman two-way analysis of variance performed on the raw (or untransformed) data yielded an χ^2_r of 8.30, statistically significant at the .05 level. This lends additional support to the finding of a statistically significant velocity effect by the parametric analysis of variance just discussed.

The nature and magnitude of the changes in average screen travel with changes in simulated aircraft speed are shown in figure 24, which plots the mean (average) values at each aircraft speed. Note that the travel for 1170 knots slightly exceeds that for 1640 knots, which is somewhat unexpected and may be a chance result. Overall, there is a trend to the data: Screen travel increases with increasing aircraft speed. Looking at the end points only, the increase amounts to .87 inches (2.6 nautical miles). This represents a 22% increase in travel with a three-fold increase in aircraft speed. Observers are responding more rapidly to false positives at higher

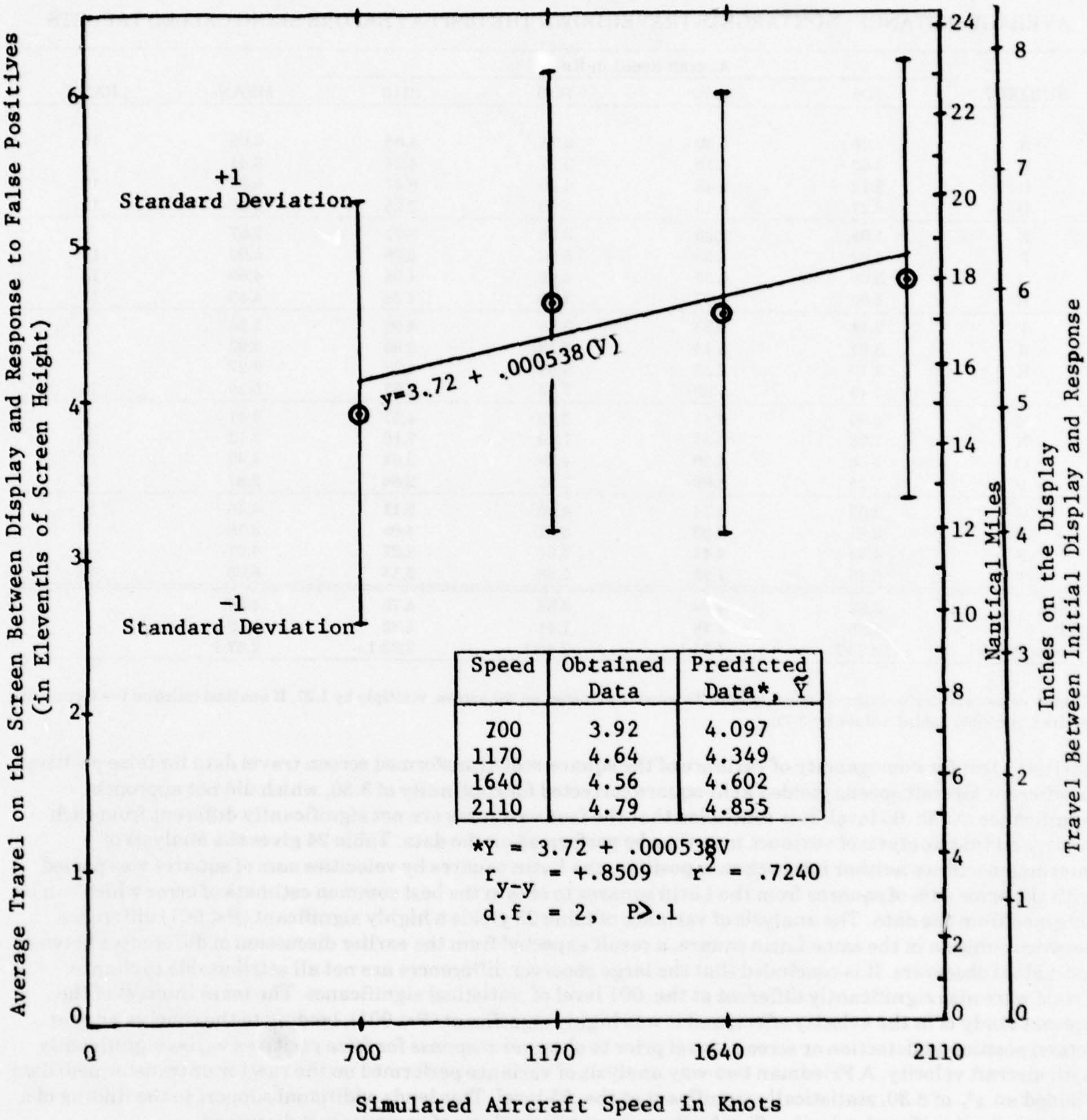


Figure 24. The interval between display and response to nontargets mistaken for targets at the four different simulated aircraft speeds.

TABLE 24

SCREEN TRAVEL+ FOR FALSE POSITIVES: ANALYSIS OF INTERACTIONS

Source of Variation	Sum of Squares	df	Mean Square	*F*
Latin Squares X Trials	.362	12	.0302	1.16
Latin Squares X Velocities	.282	12	.0235	.904
Error	.780	30	.0260	

+Square root transformed data was used.
NOTE: Neither interaction is significant at the .05 level.

TABLE 25

SCREEN TRAVEL+ FOR FALSE POSITIVES: ANALYSIS OF VARIANCE

Source of Variation	Sum of Squares	df	Mean Square	*F*
Latin Squares	.814	4	.204	7.50***
Between Subjects in Same Square	5.111	15	.341	12.5***
Velocities	.552	3	.184	6.76***
Trials	.757	3	.252	9.26***
Latin Squares X Trials	.362	12	.0302	1.11
Error (including LS X velocities)	1.142	42	.0272	

+Square root transformed data was used.
***Statistically significant at the .001 level.

simulated aircraft speeds, but they do not quite keep up with the increased aircraft speed. The nature of the distributions on the display screen, at the time observers respond, of the average number of false positives at the four aircraft speeds are portrayed in figure 25. The numbers of false positives are plotted on a logarithmic scale. For the first two elevenths of the screen the four curves are clearly separated. Farther down the screen there is considerable crossing of the curves, the slowest speed curve (700 knots) is highest in the first three intervals, but is lowest in the last six intervals where few responses are made. From the third through the eleventh interval all four curves are not far from linear in the semi-log plot, and an average of the curves would approximtely fit the equation log (N) = AT + B, where N is number detected, T is target travel or screen position, and "A" and "B" are constants. This equation may be rewritten as $N = C \times 10^{-AT}$, or as $N = De^{-FT}$ where C, A, D, and F are constants.

The plot shows that the differences established between the four curves in the first few intervals are not eliminated in later intervals. This point is clearly apparent in the cumulative frequency curves of figure 26. These curves depict the total number of responses that were made up to any given distance down the display. The plot shows that the total number of false positives increases almost linearly with screen position up to about one-third of the way down the display screen, then all four curves show an increasing tendency to flatten out. From the four curves it is clear that the slower the aircraft speed, the larger the number of false positives, and also that the increase in the number of false positives is accelerated, i.e., the spacing between the curves increases. Note the great similarity in shape and spacing between the cumulative frequency curves for target detections, figure 18, and those for number of false positives shown in figure 26.

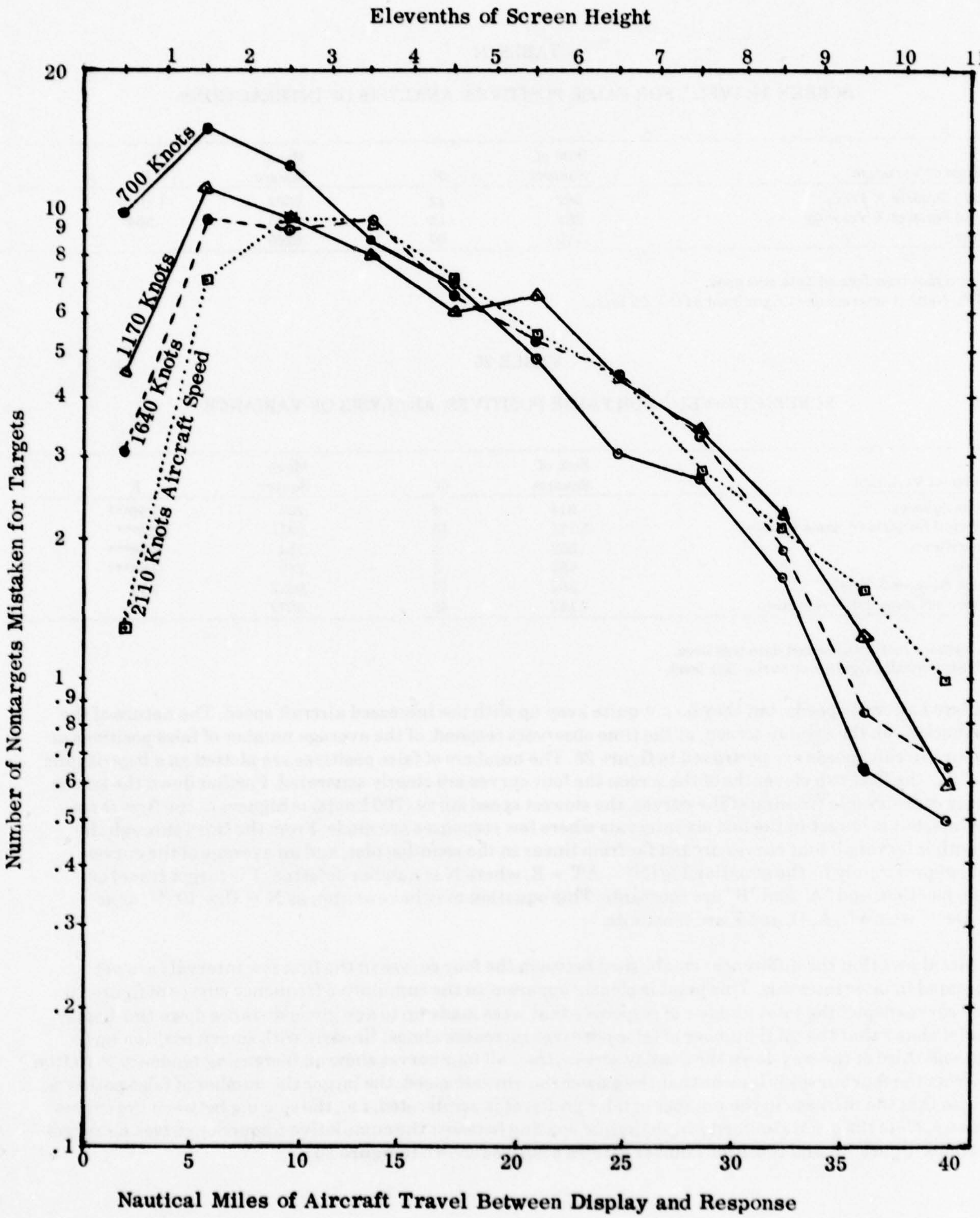


Figure 25. Number of nontargets mistaken for targets as a function of ground distance covered between the initial appearance of such objects and the observer's response to them.

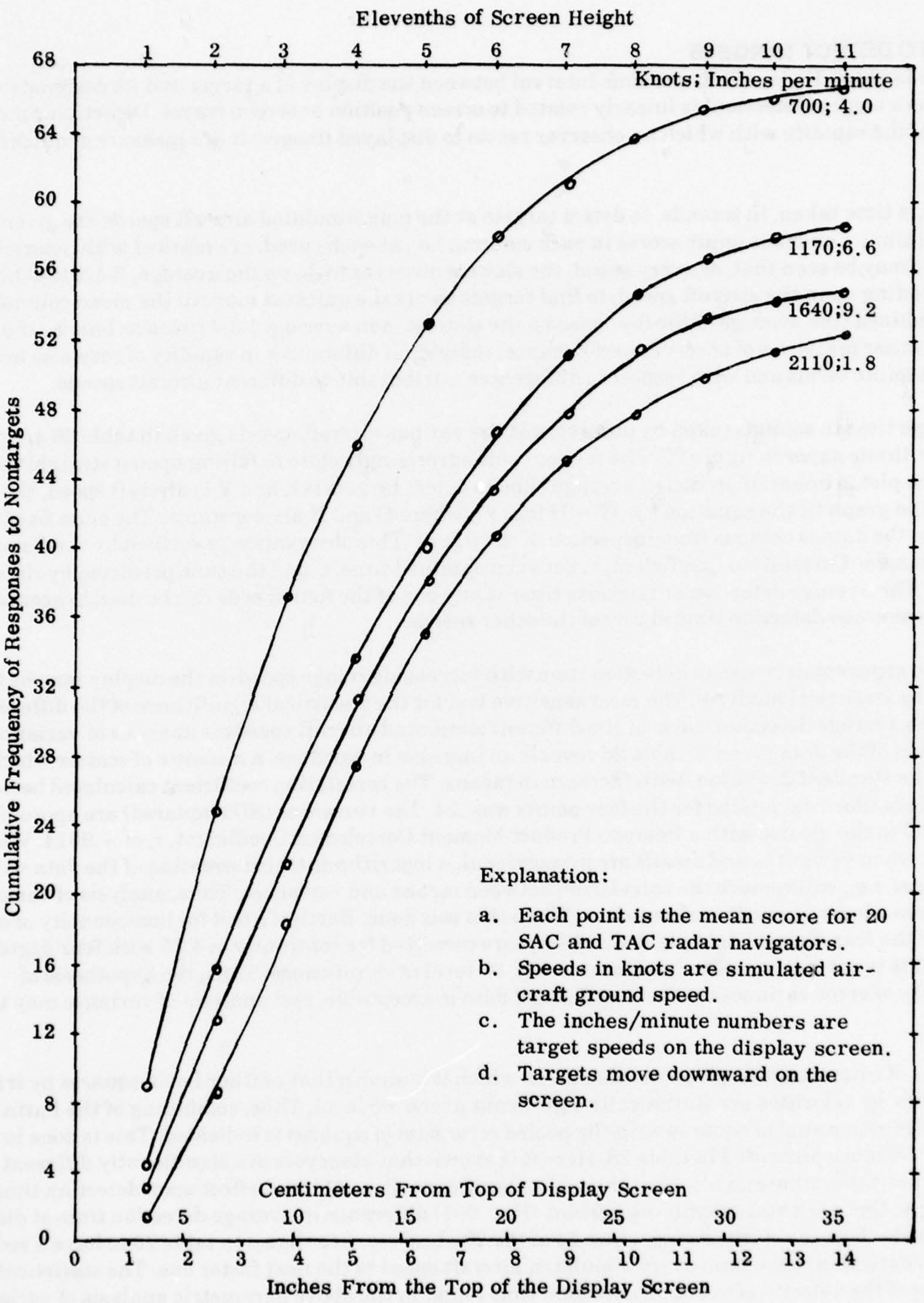


Figure 26. Cumulative Frequency of Responses to Nontargets as a Function of Screen Position for High Resolution Coherent Side-Looking Radar.

G. TIME TO DETECT TARGETS

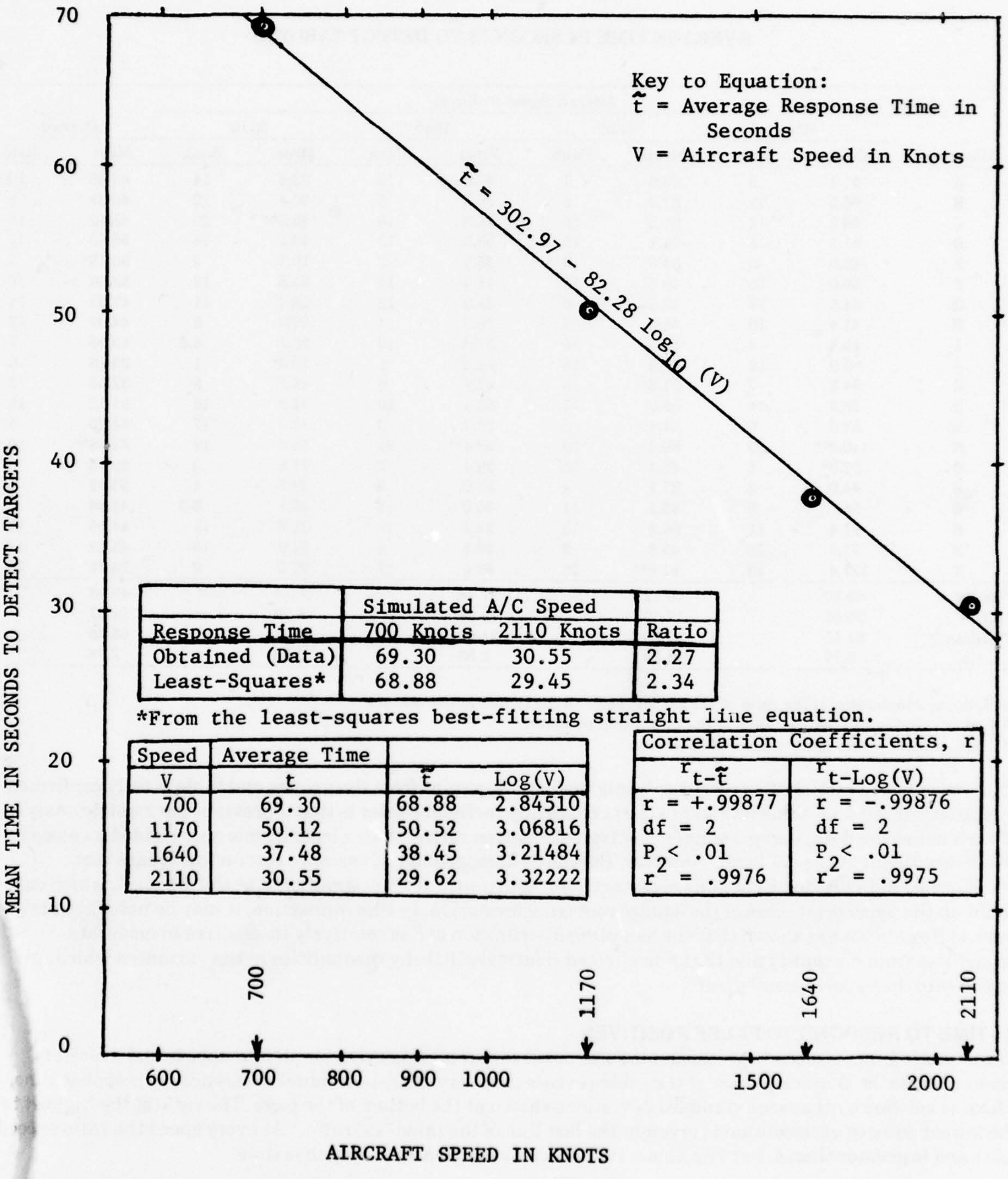
With Side-Looking Radar (SLR) the time interval between the display of a target and its designation by the observer as a target (detection) is linearly related to screen position or screen travel. Detection time is a measure of the rapidity with which an observer reacts to displayed images: it is a measure of quickness of response.

The average time taken, in seconds, to detect targets at the four simulated aircraft speeds are given in table 26. The minimum and maximum scores in each column, i.e., at each speed, are marked with asterisks. From the table it may be seen that, at every speed, the slowest observer took, on the average, 2-1/2 to 4-1/3 times as long, depending upon the aircraft speed, to find targets as did the quickest man. In the mean column for all speeds combined (the average of the four speeds) the slowest man average 2-3/4 times as long as the fastest. As with most other measures of observer performance, individual differences in rapidity of response are quite large in absolute terms and with respect to differences attributable to different aircraft speeds.

The average time in seconds taken by observers at the various aircraft speeds given in table 26 are plotted on semi-logarithmic paper in figure 27. The means come surprisingly close to falling upon a straight line, i.e., the trend in the plot is linear. If predicted average time to detect targets is $\hat{t}$, and V is aircraft speed, the data points on the graph fit the equation $\hat{t} = G - H \log (V)$, where G and H are constants. The close fit of the equation to the data is obvious from inspection of the figure. This observation is verified by the Pearson Product Moment Correlation Coefficient, r, between obtained time, t, and the time predicted by the equation, $\hat{t}$, of +.9988. The average detection or response time at any one of the four speeds can be used to accurately predict the average detection time at any of the other speeds.

The clearly apparent decrease in detection time with increasing image speed on the display may be further examined by statistical analysis. The most sensitive test for the statistical significance of the differences between the average detection times at the different simulated aircraft speeds is analysis of variance. Examination of the data given in table 26 reveals an increase in variance, a measure of scatter which is the square of the standard deviation, with increase in means. The correlation coefficient calculated between the means and standard deviations for the four points was .24. The variances (SD's squared) are approximately proportional to the means, with a Pearson Product Moment Correlation Coefficient, r, of +.9914. Winer (1962) notes that, when variances and means are proportional, a logarithmic transformation of the data will stabilize the variances, i.e., will remove the correlation between means and variances. Thus, analysis of variance of the data requires a logarithmic transformation. When this was done, Bartlett's test for homogeneity of error variance of the transformed data yielded a chi-square corrected for continuity of 4.35 with four degrees of freedom. This is not statistically significant at the .05 level of significance. Thus, the hypothesis of homogeneity of error variances of the transformed data is acceptable, and analysis of variance may be performed.

An analysis of interactions is given in table 27, in which it is shown that neither Latin squares by trials nor Latin squares by velocities are statistically significant at the .05 level. Thus, combining of the Latin squares by trials interaction sum of squares with the pooled error sum of squares is indicated. This is done in the analysis of variance presented in table 28. Here it is shown that observers are significantly different ($P < .05$) in average detection time and also that there is a significant ($P < .01$) trial effect upon detection time. More interestingly, there is a statistically significant ($P < .001$) difference in average detection time at different aircraft speeds. This is not surprising, since 14 of the 20 observers are shown in table 26 to have a reduction in reaction (or detection) time from every simulated aircraft speed to the next faster one. The statistical significance of the velocity effect upon detection time found in the above parametric analysis of variance is fully supported by the results of a Friedman Two-Way Analysis of Variance by Ranks of the original or untransformed data. This analysis yielded a χ^2_r of 40.56 with an associated probability of less than .001, i.e., a velocity effect upon detection time significant at the .001 level.



Response Time	Simulated A/C Speed 700 Knots	2110 Knots	Ratio
Obtained (Data)	69.30	30.55	2.27
Least-Squares*	68.88	29.45	2.34

*From the least-squares best-fitting straight line equation.

Speed V	Average Time t	$\tilde{t}$	Log(V)
700	69.30	68.88	2.84510
1170	50.12	50.52	3.06816
1640	37.48	38.45	3.21484
2110	30.55	29.62	3.32222

Correlation Coefficients, r	
$r_{t-\tilde{t}}$	$r_{t-\mathrm{Log}(V)}$
r = +.99877	r = -.99876
df = 2	df = 2
$P_2 < .01$	$P_2 < .01$
$r^2 = .9976$	$r^2 = .9975$

Figure 27. Average Time to Detect Targets at Various Aircraft Speeds

TABLE 26

AVERAGE TIME IN SECONDS TO DETECT TARGETS

	Aircraft Speed in Knots									
	700		1170		1640		2110		Overall	
Observer	Mean	Rank	Mean	Rank	Mean	Rank	Mean	Rank	Mean	Rank
A	51.4	5	43.6	7	47.2	15	32.5	14	43.68	10
B	66.2	12	37.9	5	26.5	5	30.4	12	40.25	8
C	76.0	14	81.2	18	48.0	16	55.3**	20	65.12	18
D	81.1	15	54.1	15	59.3	19	40.7	18	58.80	17
E	42.0	2	34.8	3	25.3	2	19.7	2	30.45*	1
F	89.0	18	49.5	12	41.4	14	34.5	16	53.60	16
G	84.5	17	42.5	6	34.3	12	29.2	11	47.62	14
H	81.4	16	45.2	8	26.3	4	23.2	5	44.02	12
I	49.1	4	54.5	16	31.6	10	25.1	6.5	40.08	7
J	60.9	10	51.9	14	24.2*	1	19.6*	1	39.15	6
K	54.3	7	21.2*	1	27.7	6	26.7	8	32.48	2
L	58.2	4	56.0	17	58.4	18	31.9	13	51.12	15
M	53.5	6	24.4	2	25.7	3	35.2	17	34.70	5
N	140.6**	20	89.4	19	62.4**	20	50.4	19	84.58**	20
O	37.3*	1	46.4	10	29.4	7	21.1	3	33.55	3
P	44.9	3	37.1	4	31.0	9	22.7	4	33.92	4
Q	55.7	8	49.1	11	36.0	13	25.1	6.5	41.48	9
R	61.4	11	50.5	13	31.7	11	32.6	15	44.05	13
S	71.1	13	45.8	9	30.5	8	27.9	10	43.82	11
T	127.4	19	91.8**	20	52.6	17	27.2	9	74.75	19
Mean	69.30		50.12		37.48		30.55		46.86	
S.D.	25.91		17.60		12.28		9.20		14.37	
Median	61.15		47.75		31.65		28.55		43.75	
Ratio+	3.77		4.33		2.58		2.82		2.78	

+Ratio, or range ratio, is the ratio of the highest score to the lowest in the column.
*Lowest score in column, i.e., time of the most rapid observer.
**Highest score in the column (slowest observer).

Thus, the faster reaction time of test subjects that was apparent from the graphs and table is fully confirmed by statistical tests. Of some interest to the statistically inclined reader is that a previous parametric analysis of variance of the data, using a square root transformation rather than a logarithmic one, yielded a velocity effect significant at the .01 level. However, Bartlett's homogeneity of variance test on the square root transformed data yielded a chi-square corrected for continuity of 23.8, significant at the .01 level, which casts doubt on the appropriateness of the square root transformation. In this connection, it may be noted that the work of Box (1953) has shown that the sampling distribution of F is relatively insensitive to moderate departures from normality and that F is affected relatively little by inequalities in the variances which are pooled into the experimental error.

H. TIME TO RESPOND TO FALSE POSITIVES

The average time in seconds to respond to false positives for all observers for all four speeds and overall is given in table 29. A quick glance at the table reveals that very large individual differences in response time which is verified by the large standard deviations shown at the bottom of the page. The ratio of the highest to the lowest score in each column is given in the last line of the table, as "ratio." At every speed the ratio exceeds 2.3:1 and is greater than 4:1 at 700 knots. Individuals vary greatly in response time.

The four means or averages at the four aircraft speeds of table 29 are plotted on semi-log paper in figure 28. Note that they fall almost on a straight line. The least-squares best-fitting line is shown on the graph. Its equation is $t = 355.94 - 98.37 \text{ Log}_{10}(V)$. The product moment correlation coefficient, r, between the average

TABLE 27

DETECTION TIME+ FOR TARGETS: ANALYSIS OF INTERACTIONS

Source of Variation	Sum of Squares	df	Mean Square	F
Latin Squares x Trials	.1876	12	.0156	1.054
Latin Squares x Velocities	.1016	12	.0085	.574
Error	.4445	30	.0148	

+ A logarithmic transformation of detection time was used.
NOTE: Neither interaction is statistically significant at the .05 level.

TABLE 28

DETECTION TIME+ FOR TARGETS: ANALYSIS OF VARIANCE

Source of Variation	Sum of Squares	df	Mean Square	F
Latin Squares	.1185	4	.0296	2.28
Between Subjects in the Same Latin Square	.4262	15	.0284	2.18*
Velocities	.4670	3	.1557	11.98**
Trials	.2241	3	.0747	5.75**
Latin Squares x Trials	.1876	12	.0156	1.20
Error++	.5461	42	.0130	
TOTAL				

+A logarithmic transformation of the data was used.
++The error of sum of squares is the pooled error sum of squares plus the Latin squares by trials interaction sum of squares.
*Significant at the .05 level.
**Significant at the .01 level.
***Significant at the .001 level.

response times and the logarithms of the aircraft speeds is −.9970, which is very high. The figure also contains a table giving obtained means or averages and means predicted from the equation. The correlation coefficient between the averages is +.9906, which is also very high. Response time decreases as the logarithm of aircraft speed. A second table gives the ratios of response time at 700 knots to response time at 2110 knots. Obtained averages show that observers responded 2.5 times as rapidly at 2110 knots as at 700 knots. The least-squares values from the prediction equation yield a ratio of 2.6. Thus, when the aircraft speed tripled, response speed increased by about 2.5 times. Speed of response did not quite keep up with the increase in speed of the aircraft. This ability to almost keep up with a big increase in speed was noted earlier for targets, where the effect was approximately the same in numerical value, i.e., 2.5.

The logarithmic equation, $t = A - B \operatorname{Log}(V)$, was also found, earlier in this paper, to also fit the data for detection time for targets. The constants A and B were different in magnitude. Some comparison of response times for targets and false positives may be obtained from examining figure 29. This figure also contains a small table. Note that the two straight lines giving the predicted response times converge. They cross over at 1960 knots, just below the maximum tested speed of 2110 knots. The value of V at the point where the lines cross each other is found by equating the right-hand sides of the two equations and solving for the value of V.

The small table included on the graph shows that observers at the 700 and 1170 knot speeds take about 11% and 7% more, respectively, on the average, to respond to false positives than to respond to targets. However, at the two highest speeds the response time differences between targets and false positives are both less than 1%. At the two lowest speeds time differences are small, and at the two higher speeds the differences are infinitesimal.

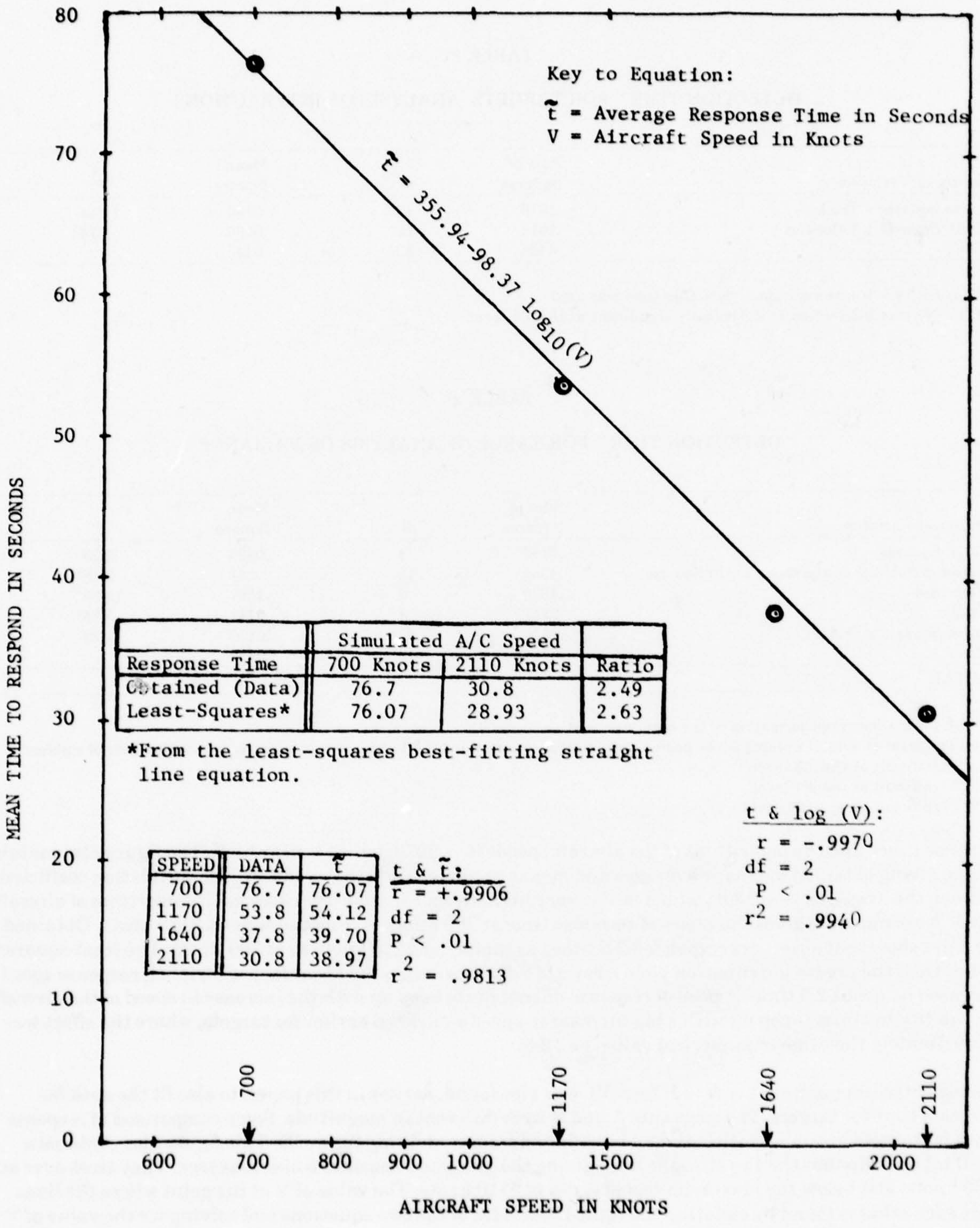
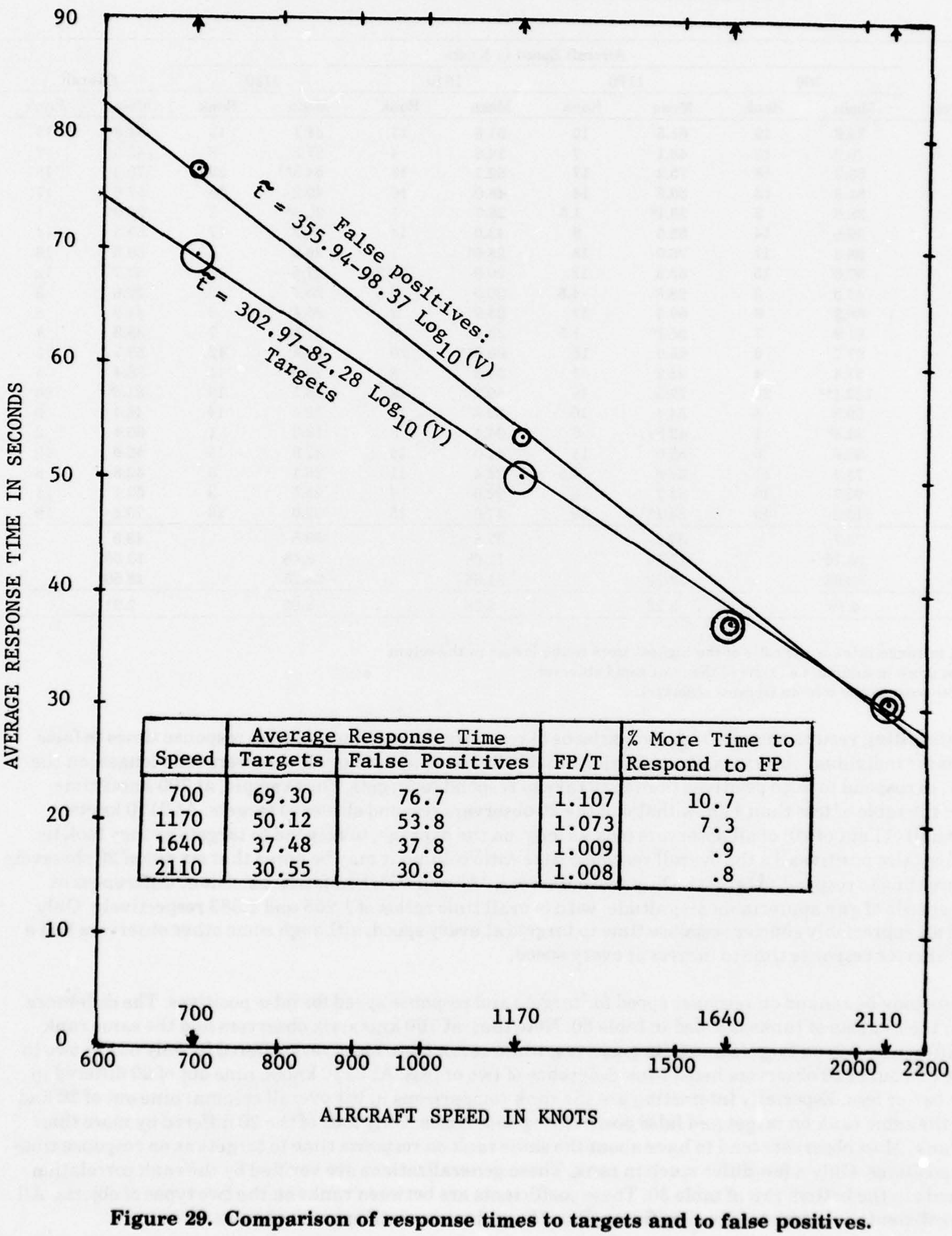


Figure 28. Average Time to Detect False Positives at Various Aircraft Speeds.



Speed	Average Response Time: Targets	Average Response Time: False Positives	FP/T	% More Time to Respond to FP
700	69.30	76.7	1.107	10.7
1170	50.12	53.8	1.073	7.3
1640	37.48	37.8	1.009	.9
2110	30.55	30.8	1.008	.8

Figure 29. Comparison of response times to targets and to false positives.

TABLE 29

AVERAGE TIME IN SECONDS TO RESPOND TO NONTARGETS

	Aircraft Speed in Knots									
	700		1170		1640		2110		Overall	
Observer	Mean	Rank	Mean	Rank	Mean	Rank	Mean	Rank	Mean	Rank
A	74.8	12	61.5	15	51.6	17	31.1	13	54.8	13
B	70.2	10	48.1	7	26.6	4	27.3	9	43.0	7
C	99.3	18	75.2	17	52.1	18	54.5**	20	70.3	18
D	82.8	13	59.5	14	48.0	16	40.2	18	57.6	17
E	38.6	2	26.1*	1.5	26.7	5	20.6*	2	28.0*	1
F	89.6	14	52.5	9	42.0	14	37.2	17	55.3	14
G	98.5	17	76.0	18	25.5*	1	26.0	4	56.5	16
H	97.0	15	55.3	12	30.9	9.5	27.5	10	52.7	12
I	47.3	3	38.6	4.5	30.9	9.5	25.7	8	35.6	3
J	68.3	9	59.5	13	25.9	2	25.4	6	44.8	8
K	61.9	7	26.1*	1.5	28.1	6	27.5	7	35.9	4
L	67.7	8	65.0	16	60.6**	20	29.4	12	55.7	15
M	57.4	4	28.2	3	30.8	8	29.0	11	36.4	5
N	142.1**	20	79.2	19	59.2	19	46.1	19	81.6**	20
O	59.3	6	54.4	10	39.6	12	32.4	14	46.4	9
P	33.9*	1	42.8	6	26.4	3	18.3	1	30.4	2
Q	58.8	5	55.0	11	41.0	13	32.9	15	46.9	10
R	73.9	11	38.6	4.5	32.4	11	26.1	5	42.8	6
S	93.7	16	51.2	8	29.9	7	25.5	3	50.1	11
T	119.2	19	84.1**	20	47.0	15	33.0	16	70.8	19
Mean	76.7		53.8		37.8		30.8		49.8	
S.D.	26.70		17.14		11.68		8.48		13.85	
Median	72.05		54.70		31.65		28.25		48.50	
Ratio+	4.19		3.22		2.38		2.65		2.91	

+ Ratio, or range ratio, is the ratio of the highest score to the lowest in the colum.
* Lowest score in column, i.e., time of the most rapid observer.
** Highest score in the column (slowest observer).

Some interesting results emerge from comparisons of response times to targets with response times to false positives for individual observers. Examination of table 30 reveals that not all observers take longer, on the average, to respond to false positives than they take to respond to targets. For example, at 700 knots time ratios in the table of less than 1 show that 4 of the 20 observers respond slower to targets. At 2110 knots, roughly half (11 out of 20) of all observers took as long, on the average, to respond to targets as they took to respond to false positives. In the overall response time ratio column it may be noted that six out of 20 observers took more time to respond to targets. Only two observers, "A" and "O", had fairly consistent differences at various speeds of any appreciable magnitude, with overall time ratios of 1.255 and 1.383 respectively. Only "O" had an appreciably shorter response time to targets at every speed, although some other observers had a slightly shorter response time to targets at every speed.

Observers may be ranked on response speed for targets and response speed for false positives. The difference between the two sets of ranks is listed in table 30. Note that, at 700 knots, six observers had the same rank (rank difference = '0) on targets and false positives, while seven more observers differed by only one or two in rank, i.e., 13 out of 20 observers had a rank difference of two or less. At 2110 knots, nine out of 20 differed in rank by two or less. Especially interesting are the rank comparisons in the overall column: nine out of 20 had exactly the same rank on target and false positive response times. Only four of the 20 differed by more than two in rank. Most observers tend to have about the same rank on response time to targets as on response time to false positives. Only a few differ much in rank. These generalizations are verified by the rank correlation coefficients in the bottom row of table 30. These coefficients are between ranks on the two types of objects. All of the coefficients are statistically significant ($P < .05$), and some are of appreciable size.

TABLE 30

COMPARISON OF RESPONSE TIMES
FOR TARGETS AND FALSE POSITIVES

	Aircraft Speed in Knots									
	700		1170		1640		2110		Overall	
Observer	Time Ratio*	Rank Diff.	Time Ratio	Rank Diff.	Time Ratio	Rank Diff.	Time Ratio	Rank Diff.	Time Ratio	Rank Diff.
A	1.46	− 7	1.41	− 8	1.09	− 2	.96	+ 1	1.255	− 3
B	1.06	+ 2	1.27	− 2	1.00	+ 1	.90	+ 3	1.068	+ 1
C	1.31	− 4	.93	+ 1	1.09	− 2	.99	0	1.080	0
D	1.02	+ 2	1.10	+ 1	.81	+ 3	.99	0	.980	0
E	.92	0	.75	+ 1.5	1.06	− 3	1.05	0	.920	0
F	1.01	+ 4	1.06	+ 3	1.01	0	1.08	− 1	1.032	+ 2
G	1.17	0	1.79	− 12	.74	+ 11	.89	+ 7	1.186	− 2
H	1.19	+ 1	1.22	− 4	1.17	− 5.5	1.19	− 5	1.197	0
I	.96	+ 1	.71	+ 11.5	.98	+ .5	1.02	− 1.5	.888	4
J	1.12	+ 1	1.15	+ 1	1.07	− 1	1.30	− 5	1.144	7
K	1.14	0	1.23	− .5	1.01	0	1.03	+ 1	1.105	− 2
L	1.16	− 4	1.16	+ 1	1.04	− 2	.92	+ 1	1.090	0
M	1.07	+ 2	1.16	− 1	1.20	− 5	.82	+ 6	1.049	0
N	1.01	0	.93	0	.95	+ 1	.91	0	.965	0
O	1.59	− 5	1.17	0	1.35	− 5	1.54	− 11	1.383	− 6
P	.76	+ 2	1.15	− 2	.85	+ 6	.81	+ 3	.896	2
Q	1.06	+ 3	1.12	0	1.14	0	1.31	− 8.5	1.131	− 1
R	1.20	0	.76	+ 8.5	1.02	0	.80	+ 10	.972	+ 7
S	1.32	− 3	1.12	+ 1	.98	+ 1	.91	+ 7	1.143	0
T	.94	0	.92	0	.89	+ 2	1.21	− 7	.947	0
Mean+	1.124	2.05	1.106	2.95	1.022	2.55	1.032	3.90	1.072	1.85
R++	.8805		.6586		.7932		.5865		.8669	

*Time Ratio = (Average response time to FP)/(Average response time to targets).

**Rank Diff. = Rank difference in response time = (Rank on targets) — (Rank on false positives).

+Mean = , i.e., average when algebraic signs are ignored.

++R = Rank correlation coefficient between rank on response time to targets and rank on response time to false positives.

NOTE: All five R values are statistically significant.

The results that have been discussed are likely a result of highly overlapping distributions of response times for targets and false positives. Any particular radar return may be responded to quickly, slowly, or not at all by different observers, or even by the same observer at different times. Also, some nontargets look more like targets than do most targets. It is clear from these observations and the discussion of response times to targets and false positives that one cannot use response time to decide if an object is a target. In other words, response time is not a criterion of response correctness. Time data do not help in solving the false positive problem with unbriefed targets. The reader may remember from the discussion of screen position or screen travel earlier in this paper that screen position or travel is a linear transform of response time, hence will not be surprised that response time, like screen position, is not useful for discriminating between targets and false positives.

I. CONFIDENCE IN RESPONSE CORRECTNESS

Observers were instructed to report, for every object that they designated as a target, how confident they were that the object was a target rather than a nontarget that looked like a target. Degree of confidence was reported by depressing the appropriate back-illuminated switch. The available choices were labeled "high confidence," "medium confidence," and "low confidence," respectively. For data reduction, high confidence was

1, medium was 2, and low was 3, respectively. Thus, the lower the number the higher the confidence.

The first question about confidence in response correctness is: "Does average confidence vary with aircraft speed?" If there is any variation, it might be that at faster speeds observers would be less confident. The data to examine for answering this question are given in tables 31 and 32, and for real target detections and for false positives, respectively. The column means, given at the bottom of the tables, do not appear to indicate any trend in either table for confidence to vary with aircraft speeds. The statistical test for an effect of speed upon confidence is analysis of variance. The analyses are given in tables 33 and 34. The small "*F*" and high value of probability (P) associated with them indicate that there is no significant velocity or speed effect. It is concluded that there is no tendency for observer confidence in response correctness to vary with aircraft speed for either real targets or false positives.

A second question about reported confidence in response correctness concerns the validity of confidence judgements: "How does confidence for real targets compare with confidence or objects mistaken for targets?" One might assume that confidence for targets would appreciably exceed that for false positives, but this may be a fallacious assumption. To make the comparison of confidence levels easier, table 35 was prepared from tables 31 and 32. It lists the ratio of average confidence for targets to that for false positives. A ratio of one means equal confidence, less than one means more confident for targets, and over one means more confident for false positives. Examination of the table shows that 6 out of the 20 observers in column 1 at 700 knots were more confident in their false positive responses than in their real target responses. Similarly, 6/20, 5/20, and 4/20 at 1170, 1640, and 2110 knots, respectively, had a greater average confidence for false positives than for real targets. This means that 20/80, or 25% of all of the observers have confidence levels counter to expectation, i.e., are less confident in responses to real targets.

TABLE 31

CONFIDENCE FOR DETECTED TARGET

Observer	Aircraft Speed in Knots 700	1170	1640	2110	Sum	Overall Average
A	1.550	1.762	1.700	1.714	6.726	1.682
B	1.500	1.769	1.733	1.920	6.922	1.730
C	2.000	1.727	1.733	2.000	7.460	1.865
D	1.259	1.242	1.222	1.429	5.152	1.288
E	1.083	1.348	1.000	1.000	4.431	1.108
F	2.235	2.133	1.788	1.454	7.600	1.900
G	1.786	1.778	1.929	1.353	6.846	1.712
H	1.467	1.450	1.769	1.611	6.297	1.574
I	2.133	2.050	1.765	1.091	7.039	1.760
J	1.667	1.684	1.113*	1.518*	5.982	1.496
K	2.143	1.588	1.652	1.538	6.921	1.730
L	2.295*	1.470	1.437	2.000	7.202	1.800
M	2.417	2.158	2.036*	2.103	8.714	2.178
N	1.417*	1.307	1.000*	1.115	4.839	1.210
O	1.240	1.222	1.364	1.421	5.247	1.312
P	1.500	1.500	1.250	1.267	5.517	1.379
Q	1.312	1.692	1.765	2.545	7.314	1.828
R	1.417	1.765	1.833	2.071	7.086	1.772
S	1.555	1.333	1.312	1.056	5.256	1.314
T	1.714	2.166	2.240	1.857	7.977	1.994
Sum	33.690	33.144	31.631	32.063	130.528	32.632
Mean	1.684	1.657	1.582	1.603	1.632**	1.632
S.D.	.391	.303	.348	.416	1.142	.285

***Due to malfunction of the data readout lamp these data were estimated.**
****Sum/80, not sun/20. The mean is 4 times the listed value.**

TABLE 32

CONFIDENCE FOR DETECTED TARGET

Observer	Aircraft Speed in Knots 700	1170	1640	2110	Sum	Overall Average
A	2.014	2.013	2.000	2.137	8.164	2.041
B	1.872	1.926	2.088	1.964	7.850	1.962
C	1.827	1.616	1.667	1.516	6.626	1.656
D	1.518	1.708	1.412	1.447	6.085	1.521
E	1.143	1.314	1.016	1.013	4.486	1.122
F	2.270	2.316	2.154	1.722	8.462	2.116
G	1.500	1.550	1.560	1.481	6.091	1.523
H	1.463	1.308	1.533	1.425	5.729	1.432
I	2.228	2.246	1.921	1.000	7.395	1.849
J	1.923	2.226	2.194*	1.924	8.267	2.067
K	2.429	1.918	2.135	2.212	8.694	2.174
L	1.568*	1.558	1.677	2.000	6.803	1.701
M	2.188	2.444	1.625*	2.362	8.619	2.155
N	1.572*	1.191	1.106*	1.281	5.150	1.288
O	1.646	1.416	1.537	1.672	6.271	1.568
P	1.346	1.542	1.357	1.432	5.677	1.419
Q	1.886	1.762	2.000	2.632	8.280	2.070
R	1.968	2.038	2.064	2.111	8.181	2.045
S	1.636	1.800	1.316	1.133	5.885	1.471
T	1.951	1.955	2.210	2.300	8.416	2.104
Sum	35.948	35.847	34.572	34.764	141.131	35.284
Mean	1.797	1.792	1.729	1.738	1.764**	1.764
S.D.	.337	.359	.373	.471	1.321	.330

*Due to malfunction of the data readout lamp these data were estimated.
**Sum/80, not sun/20. The mean is 4 times the listed value.

Although only two observers had more confidence in false positives than in targets at all four speeds, one other observer had more confidence at three speeds, two had more at two speeds, and four more had more at one speed. From a different viewpoint, in order of increasing simulated aircraft speed, 10/20, 8/20, 7/20, and 9/20, respectively, of the ratios were either in the wrong (unexpected, undesirable) direction as evinced by ratios lower than one, or else had values from .95 to 1. These ratios represent 42% of the tabled ratios. Note that in the last column, which lists the average over all aircraft speeds, only 9 of the 20 ratios exceeds .95. It is informative to examine the favorable, or numerically low, ratings. Only 2/20, 1/20, 1/20, and 1/20 of the ratios at the four speeds were .72 or lower, and all were for different observers.

The observer who had the lowest, hence most favorable, average or overall confidence ratio was "J" with a score of .730. This was achieved, in part, by obtaining an extremely good ratio of .507 at a speed of 1640 knots. However, he was not consistent, making an .867 at 700 knots. Also, of observer "J's" responses, 66.74% were to false positives. Only two other observers did slightly better on percentage of false positives. For only three observers, other than "J", was the average or overall ratio lower than .84. In the present sample of 20 radar observers one simply does not find even one who consistently averages much lower (better) confidence scores for false positives than for real targets.

From the above examination of data for individual observers it is clear that many observers are either more confident of false positive responses or else very nearly as confident of them. A comparison of group averages or means, rather than the averages of individual observers, is given in table 36. First of all, note that all five averages in the table for targets are near 1.6, with an overall mean of 1.632, while all five averages for false positives are near 1.7, with an overall average of 1.764. It is apparent for both classes of object that the average level of reported confidence in response correctness is closer to medium than to high conference.

TABLE 33

ANALYSIS OF VARIANCE FOR CONFIDENCE LEVEL FOR DETECTED TARGETS

Source of Variation	Sum of Squares	df	Mean Square	F P
Latin Squares	0.469	4	0.117	
Between Subjects in Same Square	5.724	15	0.382	
Velocities	0.171	3	0.057	.080>.90
Trials	0.244	3	0.081	
Squares × Trials	0.338	12	0.028	
Squares × Velocities	1.280	12	0.107	
Error	2.147	30	0.716	
TOTAL	10.373	79		

TABLE 34

ANALYSIS OF VARIANCE FOR CONFIDENCE LEVEL FOR FALSE POSITIVES

Source of Variation	Sum of Squares	df	Mean Square	F P
Latin Squares	2.095	4	0.524	.400>.7
Between Subjects in Same Square	6.196	15	0.413	
Velocities	0.077	3	0.026	
Trials	0.387	3	0.129	
Squares × Trials	0.362	12	0.030	
Squares × Velocities	0.489	12	0.041	
Error	1.936	30	0.065	
TOTAL	11.542	79		

From table 36 it may be seen that the average degree of confidence is greater (lower number) for targets than for false positives at every one of the four aircraft speeds, as it is for the overall mean. The statistical tests in the table for the significance of the differences in the means or averages for the two types of objects gave mixed results. At 700 knots and at 1640 knots the differences in the means did not attain statistical significance. At 1170 knots and at 2110 knots results were significant at the .05 level. For the overall, or average of the four speeds,results were significant at the .01 level, with a "t" value of 2.861, the precise value, by accident, for a probability value of .01. It may be said in summary that, although statistical significance at the .01 level was attained over the average of the four speeds, and at the .05 level for two speeds, significance was not attained at two speeds. It seems appropriate to conclude that on the average observers are more confident of target responses than false positive responses, but the differences in averages are small, barely favoring targets.

A third question about confidence is: "Are target and false positive confidence ratings correlated?" Do those who express high confidence in target responses also express high confidence in false positive responses, and similarly for low and for medium confidence observers? To answer this question, the Pearson product moment correlation coefficient for the twenty observers calculated between average confidence for target responses and average confidence for responses to false positives, using the over-all-speeds means, was +.565. This correlation, while not indicating a high degree of correlation, is statistically significant at the .01 level of statistical significance. Thus, there is a tendency for observers with above average confidence on targets to be above average in confidence on false positives, and similarly for observers with average and below average confidence levels. Clearly, different observers differ in how confident they are, yet a given observer tends to show some consistency in confidence level for the two classes of objects.

TABLE 35

(CONFIDENCE FOR TARGETS)/(CONFIDENCE FOR FALSE POSITIVES)

Observer	700	1170	1640	2110	Sum	Overall Average
	Aircraft Speed in Knots					
A	.770	.875	.850	.802	3.297	.824
B	.801	.918	.830	.978	3.527	.882
C	1.095	1.069	1.040	1.319	4.523	1.131
D	.829	.727	.865	.988	3.409	.852
E	.948	1.026	.984	.987	3.945	.986
F	.985	.921	.825	.844	3.575	.894
G	1.191	1.147	1.237	.914	4.489	1.122
H	1.003	1.109	1.154	1.131	4.397	1.099
I	.957	.913	.919	1.091	3.880	.970
J	.867	.757	.507	.789	2.920	.730
K	.882	.828	.774	.695	3.179	.795
L	1.464	.944	.857	1.000	4.265	1.066
M	1.105	.883	1.253	.890	4.131	1.033
N	.901	1.097	.904	.870	3.772	.943
O	.753	.863	.887	.850	3.353	.838
P	1.114	.973	.921	.885	3.893	.973
Q	.696	.960	.882	.967	3.505	.876
R	.720	.866	.888	.981	3.455	.864
S	.950	.741	.997	.932	3.620	.905
T	.879	1.108	1.014	.807	3.808	.952
Sum	18.910	18.725	18.588	18.720	74.943	18.735
Mean	.946	.936	.919	.936	3.747	.937
S.D.	.184	.126	.166	.138	.447	.112

TABLE 36

AVERAGE CONFIDENCE IN CORRECTNESS OF RESPONSES

Measurement	700 Knots Mean	700 Knots S.D.	1170 Knots Mean	1170 Knots S.D.	1640 Knots Mean	1640 Knots S.D.	2110 Knots Mean	2110 Knots S.D.	Overall Mean	Overall S.D.
	Aircraft Speed in Knots									
T = Target	1.686	.391	1.657	.303	1.582	.348	1.603	.416	1.632	.285
F = False Positive	1.797	.337	1.792	.359	1.729	.373	1.738	.471	1.764	.330
T/F Ratio	.937		.925		.915		.922		.925	
Student's t	1.81		2.68		2.04		2.51		2.86	
Probability*, P	>.05		<.05		>.05		<.05		=.01	
Stat. Significant	No		Yes		No		Yes		Yes	

*** The probabilty, P, is the expectation that the difference in average confidence between detected targets and false positives would be as large as or larger than the obtained difference by chance alone, provided that the true (or population) means actually were not different, given the obtained standard deviations. This P is for the obtained t with 19 degrees of freedom. The "t" values were calculated for paired observations (see Edwards, 1954, pp. 278-282).**

The confidence expressed by radar observers in the correctness of their responses in the present study may be summarized by a list of six main findings. (1) Expressed confidence did not vary with aircraft speed. (2) For all four simulated aircraft speeds and for the means of the four means, i.e., overall, average confidences for both target and nontarget objects were closer to medium confidence than to high confidence. (3) Although the average differences between the confidences for the two classes of objects were large enough to attain statistical significance at two of the speeds and for the average across speeds, all of the differences only slightly favored the targets as opposed to the false positives. (4) Fully 40% of the observers had confidence averages that were either in the wrong direction, favoring false positives, or else were negligibly different for the two classes of objects. (5) Only a few of the observers expressed much more confidence in responses to real targets than in responses to false positives, and they didn't do so at all speeds. (6) There is a low but statistically significant correlation between confidence level judgments for targets and for false positives.

From the above it may be concluded that nearly all of the observers were, on the average, almost as confident in non-target choices as in target choices. It follows that, for large unbriefed radar targets of the types used in the present study, an observer's confidence in the correctness of his designation of an object as a target has practically no value for discrimination between targets and false positives. For targets less-well resolved than those used in the present study the value of confidence judgments might be even lower. This is purely conjecture. It is clear that the confidence of observers that an object is a target is not going to be useful in solving the problem posed by an excessive number of false positives.

J. RELATIONSHIPS BETWEEN PERFORMANCE MEASURES AND SELECTION OF THE BEST OBSERVERS

A successful reconnaissance or reconnaissance/strike mission is the result of many interacting factors, not the least of which is the performance of the observer who finds the targets. In a real-time or near-real-time system, a perfect observer would find all of the targets, find them the instant that they appeared upon the display, and would not mistake any nontarget object for a target. However, the quality of the displayed picture, the target-like appearance of many images from non-target objects, and the target search process of real observers preclude perfect performance.

In addition, there are several performance measures which appear to demand somewhat contradictory behavior from the observer. The observer who tries to find all of the displayed targets cannot avoid mistaking some nontarget objects for targets: he has to be reckless. The observer who tries to avoid false positives has to be very cautious: he will miss many targets. The man who attempts to report targets almost as soon as their images appear on the display will miss many targets and report false positives: he cannot be both cautious and quick. A compromise is necessary to obtain a good overall score: Being careful but not extremely cautious, and working rapidly but not at a frantically rapid pace, etc.

The quality of an observer depends, among other things, upon the relative importance of the various performance measures. With different weightings assigned by either the test administrator or by the observers to different performance measures, different observers will turn out to be the "best." On any measure of performance, observers differ, and the performance of the same observer on any given measure fluctuates from one mission or trial to the next. Thus, selecting the "best" observers from among a group is complicated by conflicting performance measures, the relative importance of the various measures for a particular mission, instructions to the observer, and fluctuations in individuals.

If it can be shown that there is appreciable consistency or reliability in the various performance measures made on individual observers, then the data will be useful in answering the question of how to select the most efficient observers and how to avoid the least efficient. It must be kept in mind that every observer in the present study saw the same film strip four times, each time at a different simulated aircraft speed. This means there could have been some carry-over or memory effects partly responsible for trial-to-trial consistency in observer performance. However, several factors tend to minimize or largely eliminate memory effects. The side-looking radar film strip used in the present study contained the images of many targets and observers found only a small percentage of available targets. Many radar returns from nontarget objects that were imaged on the film strip could be, and indeed were, mistaken for targets. In addition, observers were tested in

other studies in which long test runs (or simulated missions) were interspersed between the trials or runs of the present study. The film strips of SLR were similar in appearance to the one used in the present study and to ones used in the training sessions. All of the many strips of film contained both many targets and many radar returns easily mistaken for targets. The performance of observers did change both absolutely and in relationship to other observers, but it is believed that memory for the specific film strip played a minor role, if any. Changes in motivation of observers, changes in interpretation of instructions, changes in target search procedure, luck or chance, etc., probably far outweighed the influence of memory for specific radar returns on the displayed imagery.

The consistency of observer responses in number of targets detected, in number of nontarget objects mistaken for targets, and in rapidity of response to targets and to nontargets is measured by the absolute magnitude or size of the correlation coefficients for the same performance measure on different trials or test runs. The data are given in table 37. Note that, in the first line of the table, performance at an aircraft speed of 700 knots is correlated with performance at 1170 knots on the same performance measures. Twenty-two, or 92% of the 24 correlation coefficients in the table, are statistically significant, 14 at the .01 level of significance. The two correlation coefficients that were too small to obtain statistical significance were both in the number of targets detected column.

The statistical significance attained by the correlation coefficients indicates that observer performance on one test run or simulated mission is somewhat predictable from performance on another test run. The amount or degree of predictability, as indicated by the size of the square of the correlation coefficients, is not high. However, it is judged that the reliability indicated by the correlation coefficients is adequate to justify examination of observer test scores for the relationships of different performance measures. These relationships are cues to the solution of the observer selection problem.

The performance of individual observers may be ranked from best to worst, a rank of 1 being the best or most desirable, while a rank of 20 is, in the present study, the worst or least desirable. Thus, the larger the number indicating rank, the poorer the performance. The ranks of individual observers allows comparison on different performance measures and is relevant to observer selection. For a first look into the selection problem, the performance or test scores of each of the 20 observers was reduced to ranks for four measures of performance and for three scores that are summed combinations of the four performance measures. The ranks of observers on each performance measure are the average over all four aircraft speeds. The data are given by table 38. In the table the best or top five scores are marked with an asterisk (*). The top and bottom scores are specially marked as an additional aid in inspecting the tabled data.

TABLE 37

OBSERVER CONSISTENCY IN NUMBER OF TARGETS DETECTED, NUMBER OF FALSE POSITIVE RESPONSES AND RESPONSE SCREEN POSITION AT VARIOUS AIRCRAFT SPEEDS

Aircraft	Product Moment Correlation Coefficients			
	Number of Responses		Response Screen Position ++	
Speeds	Targets +	False Positives	Targets	False Positives
700-1170	.6369**	.8962**	.7296*	.7826**
700-1640	.1060	.4545*	.6064**	.5242*
700-2110	.5603*	.5785**	.5135*	.6525**
1170-1640	.3081	.5356*	.7037**	.6091**
1170-2110	.5671**	.5399**	.5149*	.5780*
1640-2110	.6321**	.7804**	.6672**	.7240**

NOTE: All coefficients are based on 20 data pairs, i.e., paired scores.
+ "Number of Responses, Targets" is the number of targets detected.
++ The heading of this column could also be "Response Time," for the correlation coefficients would be the same within rounding errors.
*,** Statistically significant at the .05 and .01 levels, respectively.

TABLE 38

OBSERVER RANKINGS** ON PERFORMANCE MEASURES

Observer	Number or Percentage of Targets Detected D	Number of False Positives F	Percentage of False Positives FP	Screen Position (Speed) P	Composite Scores: Ranks Based On Sums of Ranks D+F	D+FP	D+FP+P
A	2+	18*	15	14	9.5	6.5	10
B	4+	17*	16*	7	12	10	6.5
C	19*	16*	▶ 20*	18*	▶ 20*	▶ 20*	▶ 20*
D	◆ 1+	▶ 20*	17*	17*	12	8	15
E	7	19*	19*	◆ 1+	18*	16*	6.5
F	15	2+	2+	15	3.5+	6.5	12.5
G	▶ 20*	◆ 1+	5+	13	12	14	16
H	8	5+	3+	8	◆ 1.5+	3+	2+
I	11	12	14	9	14	14	14
J	5+	10	4+	6	2+	◆ 1+	◆ 1+
K	12	6	7	2+	6	9	4+
L	6	13	9	16*	7.5	4+	10
M	3+	14	13	5+	3.5+	5+	4+
N	13	7	8	▶ 20*	9.5	11	17.5*
O	10	15	18*	3+	15.5*	17.5*	10
P	14	11	11	4+	15.5*	14	8
Q	16*	3+	6	10	7.5	12	12.5
R	17*	9	12	12	18*	19*	17.5*
S	9	4+	◆ 1+	11	◆ 1.5+	2+	4+
T	18*	8	10	19*	18*	17.5*	19*

**The performance measures (or scores) of observers are ranked (or ordered) from 1, the best score, to 20, the worst score. The ranks listed are for the average performance at all four simulated aircraft speeds. Fractional ranks, such as 2.5, or the same rank for more than one observer represent ties or equal performance. Note in the D+F column that two observers are tied with a rank of 1.5 and 3 with a rank of 12.

+The 5 best scores in the column

*The 5 worst scores in the column (six in the D+FP column due to ties).

◆ The best score in the column.

▶ The worst score in the column.

Examination of the five best scores, marked by a plus sign, in the D (Detected Targets) and F (False Positives) columns is instructive. Note that no observer with a plus in one column has a plus in the other, although observer "C" has an asterisk in both columns, indicating that an observer can be quite inferior on both performance measures. It may also be noted from the two columns that observer "D", who found and recognized the largest number of targets, mistook the most nontarget radar returns for targets. On the other hand, and in sharp contrast to this, observer "G", who detected the smallest number of targets, had the largest number of false positives. It is unlikely that, in a repeat of the present study, the very best observer on one performance measure would be the very worst on another, as was the case for these two observers. However, this occurrence is in line with, and serves to illustrate, the discussion earlier in this paper on the somewhat contradictory behavior required to maximize both performance measures.

As expected, the test scores of the remaining 18 observers, excluding observers "D" and "G", do not exhibit such a high degree of inverse or negative relationship. From table 39 it may be noted that the Spearman Rank Correlation Coefficient, r_s between rank on number or percentage of targets detected and number of false positives for the 20 observers is −.6075, which is statistically significant at the .01 level of significance. While the absolute size or magnitude of this coefficient is indicative of only a fair degree of relationship, it may be concluded that those observers who find the most targets also tend to mistake the most nontarget objects for targets, while those who find fewer targets tend to make fewer false positive responses. Doing well on either

measure tends to go with doing poorly on the other. Further examination of this tendency by reference to table 40 reveals that the relationship holds at all four aircraft speeds as well as for the overall average on the four speeds.

Referring back to table 38, it may be noted, from inspection of the D and the FP columns, that the observer ranks on the percentage of responses that are made to nontargets does not have any noticeable relationship to the number of targets detected. This observation is confirmed by table 40, which gives the rank correlation coefficient, r_s, between either the number or the percentage of available targets detected and the percentage of false positives. For all speeds combined the coefficient is only −.2241. This value is too small to be statistically significant, indicating that the data contain no statistically valid evidence for other than a chance relationship. This finding is borne out by the data of table 41, which show that the coefficient of correlation is not larger than would be expected by chance causation alone at every one of the four simulated aircraft speeds as well as for the overall average of the four speeds. It may be concluded that, in the present study, there is no

TABLE 39

CORRECTIONS+ BETWEEN OBSERVER RANKS ON FOUR PERFORMANCE MEASURES OR SCORES++

Performance Measure or Observer Score	P = Screen Position When Response Occurred	FP = Percentage of False Positives	F = Number of False Positives
D = Number or Percentage of Targets Detected	+.2737	−.2241	−.6075**
F = Number of False Positives	+.0447	+.8827**	—
FP = Percentage of False Positives	−.0902	—	—

+The correlation coefficients in this table are Spearman Rank Correlation coefficients, each based on 20 data pairs or observer scores.

++The scores or measures used are the average or mean rank of individual observers over the 4 simulated aircraft speeds.

**Statistically significant at the .01 level of significance: The correlated variables are related, i.e., either is, to some extent, predictable from the other.

TABLE 40

CORRELATION BETWEEN THE NUMBER OR THE PERCENTAGE OF AVAILABLE TARGETS DETECTED AND THE NUMBER OF NONTARGETS MISTAKEN FOR TARGETS+

Aircraft Speed in Knots	Correlation Coefficient++	
	Product-Moment, r	Rank r_s
700	.7509**	−.5559**
1170	.6417**	−.4739*
1640	.5577**	−.3632
2110	.6827**	−.5443**
Overall	.6732**	−.6075**

*, **Significant at the .05 and .01 levels of statistical significance, respectively, by a one-tailed test of significance.

+Number of nontargets mistaken for targets is the same as number of false positives.

++The two correlation coefficients are the Pearson Product Moment, r, and the Spearman Rank Correlation Coefficient, r_s. The r_s values have been corrected for ties in ranks. Since low numberical values (high ranking) go with high numbers of detected targets and low numbers of false positives, the r and r_s coefficients are opposite in algebraic sign.

statistically valid evidence that either the number or the percentage of available targets detected is related to the percentage of all responses that are made to nontarget radar returns or to its linear transform, accuracy. Accuracy is the percentage of observer responses that are made to genuine targets. In different phraseology, observers who excell in finding many targets are just as likely to be poor at accuracy or percentage correct or at percentage of false positives as are observers who are mediocre or poor at detecting many targets. Since the performances are unrelated, i.e., neither is predictable from the other, if both are important, then observer selection requires measurement of both.

The percentage of false positives is defined as 100 times (number of false positives)/(number of false positives + number of targets detected). Thus, it would not be surprising to find that the number and the percentage of false positives are correlated scores. Inspection of observer ranks in the number of false positives (F) column and the false positive percentage (FP) column of table 38 reveals that low numbers (good performance) in either column tend to go with low numbers in the other, and the same relationship holds for medium sized numbers and for large numbers (poor performance). This relationship is supported by table 42, which shows that the number of false positives is highly related to the percentage of false positives: $r_s = .8827$,

TABLE 41

CORRELATION BETWEEN NUMBER OR PERCENTAGE OF AVAILABLE TARGETS DETECTED AND THE PERCENTAGE OF RESPONSES THAT ARE FALSE POSITIVES

Aircraft Speed in Knots	Correlation Coefficents+	
	Product Moment, r	Rank, r_s
700	+.3386	−.2896
1170	+.1686	−.1137
1640	−.0352	−.0489
2110	+.0150	+.0241
Overall	+.2597	−.2241

+ The correlation coefficients are Pearson Product Moment, r, and Spearman Rank, r_s, each based on 20 data pairs or paired scores.

NOTE 1. None of the tabled coefficients are large enough to attain statistical significance at the .05 level of significance: No relationship has been shown to be present between the correlated variables.

NOTE 2. Since, by definition, percentage accuracy = 100−(percentage of false positives), it follows that, except for a change in algebraic sign, the correlation coefficients tabled above measure the relationship between response accuracy and number or percentage of available targets detected. Hence, there is no evidence for a relationship between accuracy and number or percentage of targets detected: the variables are independent of each other.

TABLE 42

CORRELATION BETWEEN NUMBER OF FALSE POSITIVES AND PERCENTAGE OF FALSE POSITIVES FOR INDIVIDUAL OBSERVERS

Aircraft Speed in Knots	Correlation Coefficient	
	Product Moment, r	Rank, r_s
700	.8228**	.8612**
1170	.7933**	.8902**
1640	.7689**	.8597**
2110	.7540**	.7726**
Overall+	.8549**	.8827**

**Statistically significant at the .01 level of significance.

All coefficients, including the "overall" r's, are based on 20 data pairs or 20 scores.

+ Overall = correlations based on the averages of the 4 scores for the four aircraft speeds for each observer.

NOTE: The statistical significance of the correlation coefficients means that the percentage of false positives, or its inverse accuracy, are related to or predictable from the number of responses made to nontarget radar returns.

a positive correlation which is large enough to indicate a strong degree of relationship and to attain statistical significance at the .01 level of significance. Table 42 lists both product moment and rank correlation coefficients at all four aircraft speeds and over all speeds between number of false positives and percentage of false positives. All of the tabled coefficients are of appreciable size, i.e., .75 or larger, and all are statistically significant at the .01 level of significance, verifying the relationship noted between the variables. It is clear that doing well (or doing badly) on number of false positives goes along with doing well (or doing badly) on percentage of false positives. Performance on either measure is, to an appreciable extent, predictable from performance on the other.

If action is to be taken against a target, it may be necessary to recognize and designate target images very soon after their appearance upon the display, i.e., quick responses may be important. How far a target image has moved down the display before it is responded to by the observer, called screen position or display position, is a measure of quickness of response.

Examination of the screen position, or P, column of table 38, the observer ranking table, reveals that the best observer (Rank 1) in the column ranked 19, next to worst, in both the F and FP columns. In contrast, the second best man in the column did well with a six and seven in these columns. The two worst observers in the column were mediocre on these measures. More extensive comparisons of the P column rankings with those in the D, F, and FP columns reveal no apparent relationship between screen position rankings and rankings in these three columns. The first, or P, column of table 43 reveals that all three of the correlation coefficients are too small to attain statistical significance, verifying the observed lack of relationship. This is seen to hold true over every simulated aircraft speed for both number of detected targets and false positives in table 44. Since numbers of targets detected, numbers of nontargets mistaken for targets and percentage of false positives, or its inverse, accuracy, are not useful measures in selecting observers for short reaction times, it follows that quickness must be assessed independently of them.

A question of some interest to observer evaluation is "are some observers good to excellent on most or all performance measures and are some mediocre to inferior on most or all measures?" Inspection of the observer ranking table (table 38) reveals that no observer had a rank of better (lower) than five on all four position measures. However, observer "H" was eight or better on all four, and observers "J", "K" and "H" were eight or better on three out of the four scores in the table, only observers "J", "K", "H" and "S" had ranks of 12 or better

TABLE 43

CORRELATION BETWEEN NUMBER OF RESPONSES AND AVERAGE DISTANCE DOWN THE DISPLAY AT WHICH RESPONSES WERE MADE

Aircraft Speed in Knots	Correlation Coefficients			
	Detected Targets		False Positives	
	Product Moment, r	Rank, r_s	Product Moment, r	Rank, r_s
700	−.3533	+.3820	−.2309	−.2535
1170	−.2278	+.1733	−.2470	−.3816
1640	−.1801	+.3607	−.0316	−.0444
2110	−.0383	−.0105	+.0878	+.1175
Overall	−.1951	+.2737	+.0158	+.0447

None of the tabled coefficients are statistically significant: Number of targets detected appears to be unrelated to the rapidity with which they are detected, and the number of false positives appears to be unrelated to the rapidity with which responses are made to them. All of the tabled coefficients are based on 20 pairs of scores and the rank correlation coefficients are corrected for ties in ranks. Since detecting more targets is a better performance and yields a smaller number for rank, the product moment and rank correlation coefficients are opposite in algebraic sign except when both are close to zero, in which case the algebraic signs may either agree or disagree.

TABLE 44

CORRELATION BETWEEN PERCENTAGE OF FALSE POSITIVES AND AVERAGE DISTANCE DOWN THE DISPLAY AT WHICH RESPONSES WERE MADE

Aircraft Speed in Knots	Correlation Coefficient, r_s	
	Detected Targets	False Positives
700	−.3932	−.1872
1170	−.1639	−.2458
1640	−.0226	+.0767
2110	+.1501	+.2737
Overall	−.0752	−.0617

All correlation coefficients are Spearman Rank Correlation Coefficients, and all are calculated from 20 paired observer scores. None of the coefficients are large enough to attain statistical significance; there appears to be no relationship between percentage of false positives and the average rapidity with which observers respond to either targets or to objects mistaken for targets.

(lower number) on all four scores. If a rank of 15 or worse is "bad", then only observer "C" was bad on all four scores and observers "C" and "D" were bad on three out of four scores. However, seven observers (B, C, D, E, F, O, T) had half or more (two or more) of their scores that were 15 or worse. It appears, then, that an observer may be found who rates good to excellent on all performance measures, though not excellent on all, and that an observer can be found who is decidedly inferior on all or most all of the performance measures discussed up to this point.

Since any situation or mission may have its own overall goal, the relative importance or weight assigned to any particular performance measure will vary with the mission. Since no observer in the present study had excellent scores on all measures, the observer selection question is complex. In most situations it is likely that more than one performance measure will be important. Composite scores made by combining scores can be formed in countless ways, and in each combinatory procedure the weight of each score relative to the others can vary. A complex approach would be, for example, to convert scores to standard deviations from the mean and then average these. A simple approach would be to simply add ranks or multiply ranks and rerank the results. Thus, ranks on numbers of targets detected could be added to ranks on number of false positives. The high correlation of the scores on these two measures means that the composite would not be much of an improvement on either one used alone. A preferable method would be to add detection rank to percentage of false positive rank, since these scores are not correlated. A third method would be to combine detection, percentage of false positives and screen position ranks. Of course, all four of the primary scores in the ranking table could be added. The first three of these simple rank addition scores are given in table 38. In the composite scores columns it may be noted that, as expected, the "best" observer is a different person in each composite, although the worst is not. Four observers have plus signs attached to their scores in all three columns, indicating that they are among the five best observers in each column. Three observers have asterisks in all three columns, indicating that they are among the five worst in all three columns. It is clear that generally good observers may be selected and generally bad observers rejected using composite scores. This is possible even though best for tasks varies with the task, so that the best man by one measure or with one set of instructions is not necessarily the best by another.

The main findings of this section may be summarized as follows: (1) observers who find a large percentage of available targets also mistake a large number of nontargets for targets. The correlation of the measures means that, in general, doing well on either one goes along with doing poorly on the other; (2) there is no indication in the data that number of targets detected is related to percentage of false positives or to its linear transform, accuracy; (3) there is a high and positive relationship between the number of false positives and the percentage of responses that are false positives; (4) the target detection time score, or mean screen position at which targets are found and designated, is not related to the numbers of targets detected, to the numbers of nontargets mistaken for targets, or to the percentage of responses that are false positives or its inverse,

accuracy; (5) despite the somewhat contradictory behavior requirements for doing excellently on different measures of performance, there is a small percentage of observers who do well on all or almost all common performance measures. Similarly, a few observers are inferior on all or almost all measures. Some people are superior observers and some are inferior, even over a range of performance measures.

K. RADAR RETURNS AND THE FALSE POSITIVE PROBLEM

It was shown earlier in the present report that the percentage of available targets that were detected by the average observer at any simulated aircraft speed was low. In addition, it was found that false positives, that is nontargets identified by observers as targets, considerably exceeded the number of detected targets. Both results are undesirable. To gain some insight into what was responsible for the poor performance, every one of the hundreds of responses made by observers at the simulated speed of 700 knots was examined. This was done by projecting the data camera pictures onto a screen. The pictures showed what was on the display when an observer reported the presence of a target. Also shown was his pointer or wand indicating what radar return was designated as a target. Every radar return identified by any observer was examined on the screen and the number of individuals identifying that return as a target was tabulated. This was facilitated by marking all radar returns that were responded to by one or more observers with a marker pen on a copy of the five-inch-wide radar film viewed directly on a light table. The tabulated data are given in table 45 and are plotted on a graph in figure 30.

Note that the graph shows the number of targets responded to by one person, responded to by two persons, etc. The data are given in a similar fashion for nontargets mistaken for targets. The graph also depicts the ratio of false positives to targets. The probability that a radar return will be identified (responded to) as a target is defined as the relative frequency of response to it, i.e., the proportion of observers who called it a target. A few examples will clarify the meaning of the data in the table. From the column headed by "T Targets" it can be seen that 18 of the 78 targets were not detected by even one of the 20 observers, eight were detected by one observer, only one target was detected by all 20 observers, etc. Similarly, the table shows that 160 nontarget radar returns were identified as targets by only one observer, while each of five nontarget returns were called targets by 14 different observers, etc.

From the table it may be seen that only 60 (8 + 8 + ... 1) of the 78 targets that had been prejudged by the experimenter and co-workers as detectable were detected by one or more people, i.e., had a detection probability at a 700-knot aircraft speed of .05 or higher. When the number of people who detected each target is summed across targets and then divided by the number of targets, the result is 178/60 = 2.97. Thus, the average number of observers who deteted each target was 2.97. This value is very close to 3, the nearest integer.

It is interesting to examine the number of nontarget radar returns that are also responded to by 3 or more observers. From the table it can be seen that there were 43 + 25 + . . . + 5 = 160 of thém. Comparing this to the 60 targets detected by 3 or more observers yields 160/60 = 2.67 times as many nontargets as targets that had a "detection" probability equal to that of the average target.

From inspection of the table and the graph, the following additional facts are apparent:

1. For all detection probabilities less than .5, the number of false positives greatly exceeds the number of detected targets. These nontarget radar returns with absolutely (but not relatively) low detection probabilities account for most of the false positives.

2. For probabilities of response of .5 through .7, the number of false positives equals or exceeds the number of detected targets.

3. No radar return from a nontarget has a detection probability exceeding .7. However, only six of the 78 targets that were judged as detectable had a detection probability over .7, i.e., less than eight per cent of the targets had a detection probability of .7 or greater.

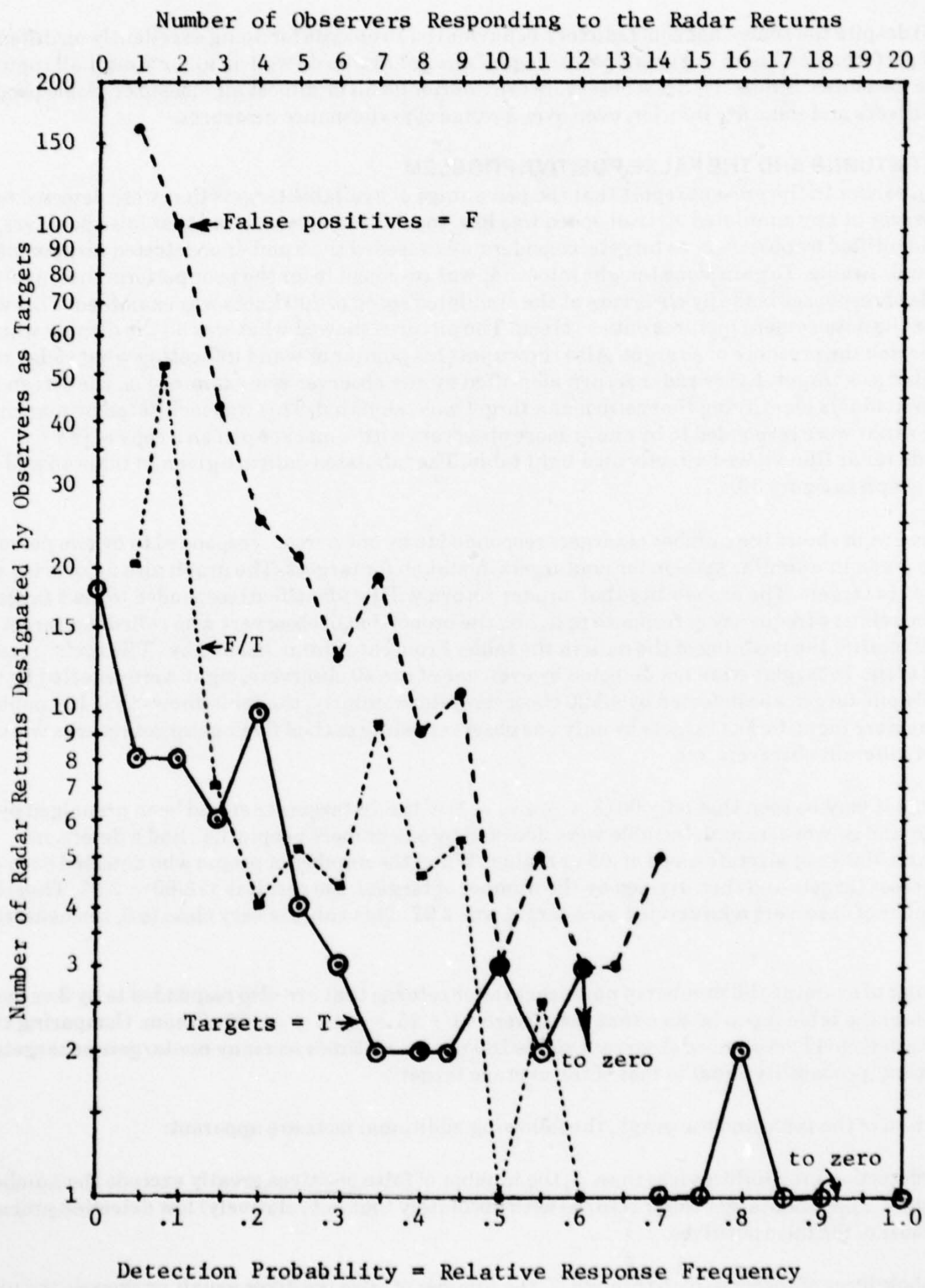


Figure 30. Numbers of Targets and Nontarget Objects at Various Observer Response Frequencies and the Ratio of the Numbers of the Two Types of Responses.

The high frequency of nontarget radar returns responded to as targets by several subjects, as indicated by the preceding discussion and an examination of the false positives/targets, or F/T, curve on the graphs is highly significant. The inference to be drawn is that many false positives are not entirely the product of active or overactive observer imaginations. The popularity of such radar returns appears to be due to their great resemblance to targets. The "popular" nontargets marked on the radar film during the data collection for the table already discussed in this section were examined. It was found, as expected, that most of these popular nontarget returns looked very much like targets. In a word, most of the nontargets identified by several observers as targets had target signatures not distinguishable by the examiners from the returns of real targets.

The element of imagination, in contrast, may be responsible for an appreciable portion of the nontargets mistaken for targets by only one or two observers. If this is the case, then more extensive training with heavy emphasis on reduction in the number of unpopular false positives may be of some value in reducing the relative frequency of their occurrence. However, it appears likely that more training can do little or nothing to reduce the large numbers of the more popular false positives without a high cost in terms of detected real targets. With more training, it is quite probable that the most popular nontarget radar return will continue to look more like real targets than do most unpopular real targets. Possibly equipment techniques, such as multisensors using different portions of the electromagnetic spectrum, can be of value in reducing the magnitude of the false positive problem.

Returning to the data, it will be noted from an examination of the semilogarithmic plot of figure 30 that, in the range of 1-10 or so observers, if fluctuations in the data are ignored, both the target and the false positive data would not deviate much from straight lines. This linearity of the semilog plots means that the two curves may be described by equations of the form $\log(N) = A - B(n)$, where N is number of radar returns, n is number of observers or detection probability and A and B are constants. When solved for N explicitly, this logarithmic equation yields the exponential equation $N = 10^{A - Bn}$, or $N = e^{C - Dn}$.

The fluctuations in the data for both targets and nontargets are sufficiently large to raise the question of the closeness of fit of the data to straight lines. These fluctuations may be largely smoothed out by plotting cumulative frequencies rather than frequencies. This was done using the data in table 45, yielding figure 31. The straight lines in the figure are the least-squares best fits to the data, using 16 data points ($n = 1 - 16$) for targets and nine data points ($n = 5 - 14$) for false positives. The fit of the cumulative frequencies, Σ N, to the straight lines is clear upon examination of the graph. The closeness of the fit is indicated by a Pearson product moment correlation coefficient, r, between obtained data and value predicted by the equation, of .9903 for targets and .9948 for false positives.

When best-fit equations are derived for $n = 1 - 20$ observers for targets and $n = 1 - 14$ for false positives (for $n > 14$, $N = 0$ for false positives), the constants in the equations change somewhat, and the correlations, as expected, drop. The correlation coefficients relating obtained and equation-predicted values for this more extended range then become $r = .9420$ for targets and $r = .9885$ for nontargets.

While use of product moment correlation coefficients relating predicted and obtained frequencies yields some insight into the degree of relationship, the proper statistic for testing goodness of fit of the data to the derived exponential equations in chi-square. Since this is defined to be $\Sigma\,[(X_T - X)^2/X_T]$, where X_T is the theoretical (or equation-supplied) value and X is the obtained value, the smaller the value of chi-square the better the fit of the equation to the data. Table 46 indicates the fit of the data to equations derived for various ranges of n, number of observers.

From the chi-square table it is clear that targets for n of $1 - 16$, the exponential equation is an excellent fit, but when $n = 16 - 20$ is included, the fit is not good. On the other hand, the fit for the nontargets is excellent when the low values of n of $1 - 4$ is omitted, fair if only one is omitted, and poor if $n = 1$ is included.

TABLE 45

NUMBERS AND CUMULATIVE NUMBERS OF OBSERVERS RESPONDING TO TARGET AND TO NONTARGET RADAR RETURNS AT 700 KNOTS

n^{+} Observers	P^{++}	T Targets	F = False Positives	F/T	Cumulative* Distribution n	T	F	F/T
0	0	18				78		
1	.05	8	160	20.0	1-20	60	421	7.02
2	.10	8	101	50.5	2-20	52	261	5.02
3	.15	6	43	7.2	3-20	44	160	3.64
4	.20	10	25	2.5	4-20	38	117	3.08
5	.25	4	21	5.2	5-20	28	92	3.29
6	.30	3	13	4.3	6-20	24	71	2.96
7	.35	2	19	9.5	7-20	21	58	2.76
8	.40	2	9	4.5	8-20	19	39	2.05
9	.45	2	11	5.5	9-20	17	30	1.76
10	.50	3	3	1.0	10-20	15	19	1.27
11	.55	2	5	2.5	11-20	12	16	1.33
12	.60	3	3	1.0	12-20	10	11	1.10
13	.65	0	3	3/0	13-20	7	8	1.14
14	.70	1	5	5/0	14-20	7	5	1.40
15	.75	1	0	0/1	15-20	6	0	0
16	.80	2	0	0/2	16-20	5	0	0
17	.85	1	0	0/1	17-20	3	0	0
18	.90	1	0	0/1	18-20	2	0	0
19	.95	0	0	0/1	19-20	1	0	0
20	1.00	1	0	0/1	20	1	0	0

*Cumulative = Additive, e.g., 60 targets were detected by from one to twenty observers, i.e., by one or more observers, 21 by 7 or more etc.

$^{+}$n = Number of observers who responded to the radar return

$^{++}$P = Response Probability = Relative Response Frequency = n/20

TABLE 46

CHI-SQUARE TESTS OF GOODNESS OF FIT OF THE DATA TO EXPONENTIAL EQUATIONS

	Values of n: TARGETS		NONTARGET RADAR RETURNS		
	1-20	1-16	1-14	2-14	5-14
Chi-Square	30.984	1.096	20.086	7.050	2.000
df*	19	15	13	15	9
P**	<.01	>.99	<.01	>.80	>.99

*df = Degrees of freedom = (number of values) − 1

**P = Probability of as large a value of chi-squares due to sampling (chance) if the true or population deviation from a perfect fit is zero.

Note from figure 31 that the best-fit equations for both targets and nontargets have the form $\Sigma\tilde{N} = A^{A-Bn}$. Taking $\tilde{N}$ to the continuous case, rather than the discrete form implied by the summation sign, yields $\int\tilde{N}dN = 10^{A-Bn}$. When both sides of the equation are differentiated, the result is $\tilde{N} = (10^{A-Bn})(\log_e 10)(-B) = (-2.303B)(10^{A-Bn})$. The constant $-2.303B$ may be replaced by 10^E to yield $\tilde{N} = 10^{(A-Bn)+E} = 10^{(A+E)-Bn} = 10^{D-Bn}$, which is of the same form as $\tilde{N} = 10^{A-Bn}$, or as $\tilde{N} = e^{A-Bn}$. Thus, the simple exponential equation which was said to describe the data from inspection of the curves of figure 30 is adequate. Note, however, that the n here is cumulative, i.e., one or more, two or more, etc.

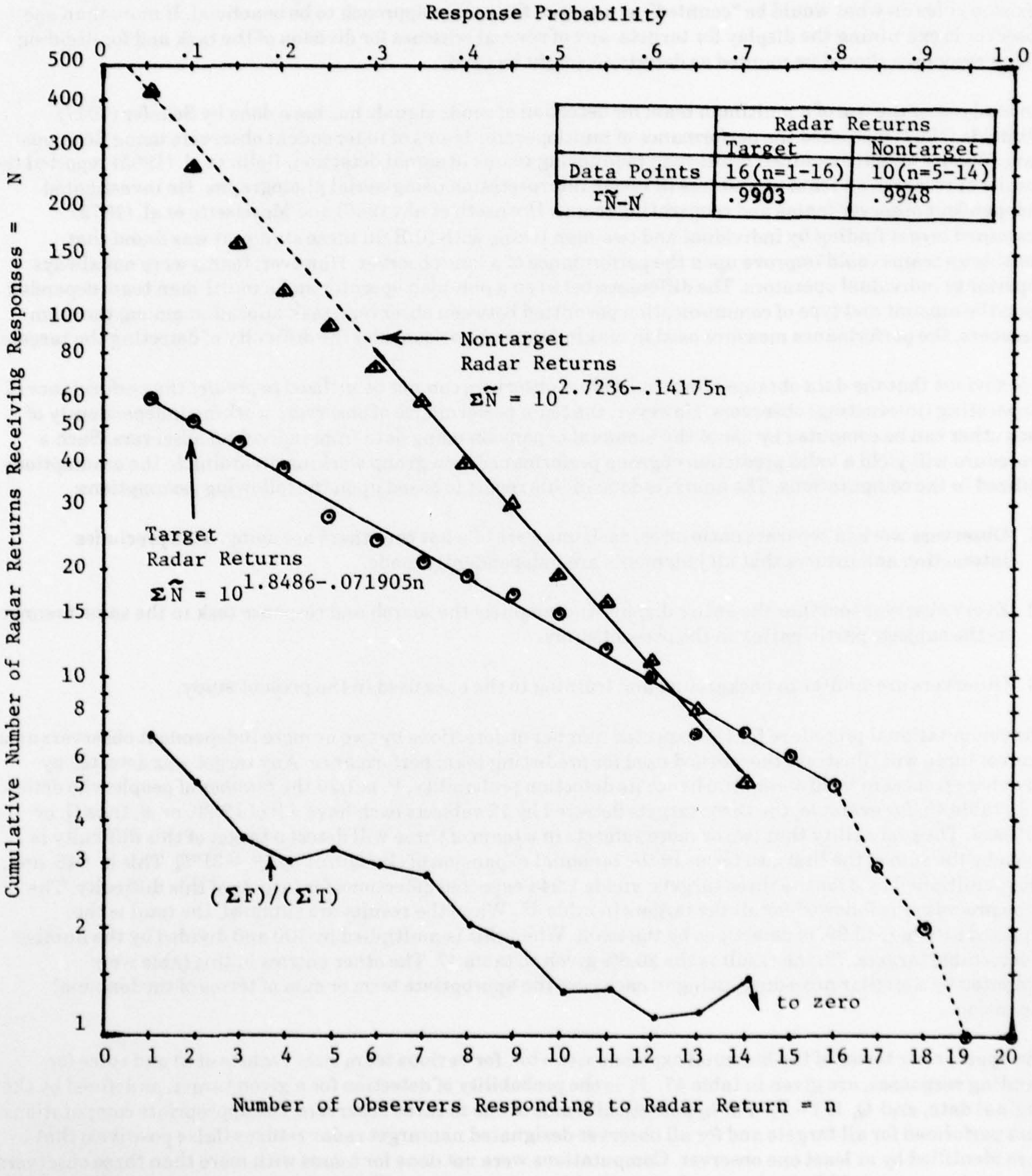


Figure 31. Cumulative numbers of target and nontarget returns, Σ N, designated by n or more observers and the ratio of the cumulative numbers of nontarget to target responses.

L. UTILIZATION OF TEAMS OF INDEPENDENT OBSERVERS

The large number of targets with low detection probabilities and the presence of many nontargets with low probabilities of being mistaken for targets present a situation in which one might expect, with the optimum decision rules on what would be "counted" as a target, for a team approach to be beneficial. If more than one observer is examining the display for targets, any of several schemes for division of the task and for deciding which responses should be counted as detections might be used.

An analysis of the use of a multiman team for detection of sonar signals has been done by Schafer (1947). Whitside (1957) discussed the performance of multioperator teams of independent observers using fictitious data. Wiener (1964) examined multi-man monitoring teams in signal detection. Bolin et al. (1965) reported the results of research on team procedures in image interpretation using aerial photographs. He investigated independent observer teams and cooperating teams. Hornseth et al. (1966) and Morrisette et al. (1975) examined target finding by individual and two-man teams with SLR. In these studies it was found that multi-man teams could improve upon the performance of a lone observer. However, teams were not always superior to individual operators. The difference between a one-man operator and a multi-man team depended upon the amount and type of communication permitted between observers, task allocation among the team members, the performance measure used in making the comparison, and the difficulty of detecting the targets.

It is obvious that the data obtained from individual observers can not be utilized to predict the performance of cooperating (interacting) observers. However, the team performance of observers working independently of each other can be computed by use of the binomial expansion using data from individual observers. Such a procedure will yield a valid prediction of group performance for a group working according to the assumptions utilized in the computations. The analysis done in this report is based upon the following assumptions:

1. *Observers work in separate enclosures, each unaware of what the others are doing. This precludes* interaction and insures that all judgments are independently made.

2. Every observer searches the entire display and conducts the search and response task in the same manner as the subjects participating in the present study.

3. Observers are similar in background and training to the ones used in the present study.

The computational procedure for the expected number of detections by two or more independent observers on a team of three will illustrate the method used for predicting team performance. Any target was detected by anywhere from zero to 20 observers, hence its detection probability, P, is 1/20 the number of people who detect it. In table 45, for example, the three targets detected by 12 subjects each have a P of 12/20, or .6, thus Q, or 1-P, is .4. The probability that two or more subjects in a team of three will detect a target of this difficulty is given by the sum of the first two terms in the binomial expansion of $(P+Q)^3$, or by $P^3 + 3P^2Q$. This is .648, and, when multiplied by 3 for the three targets, yields 1.944 expected detections for targets of this difficulty. The same procedure is followed for all the targets in table 45. When the results are summed, the total is the expected number, 15.99, of detections by the team. When this is multiplied by 100 and divided by the number of detectable targets, 78, the result is the 20.5% given in table 47. The other entries in this table were computed by a similar procedure, using in each case the appropriate term or sum of terms of the binomial expansion.

The appropriate terms of the binomial expansion, $(a+b)^n$, for various team sizes (values of n) and rules for counting responses, are given in table 47. P_i is the probability of detection for a given target, as defined by the original data, and Q_i is $1-P_i$. For hypothetical teams of one to three observers, the appropriate computations were performed for all targets and for all observer designated nontarget radar returns (false positives) that were identified by at least one observer. Computations were not done for teams with more than three observers since such teams would be too large to be practical in most situations. Table 48 summarizes the results of the computations, giving the percentage of the targets detected and the percentage of false positives for independent observer teams of various sizes with various decision rules on what should be "counted" as a target.

TABLE 47

COMPUTATION TERMS FOR VARIANCE DECISION RULES

Number of Observers	Decision Rule: include all responses made by:	Quantity for Computing Group Detection Probability
1	The Observer	P_1 (the original data)
2	Both Observers	P_1^2
2	Either or Both Observer(s)	$P_1^2 + 2P_1Q_1$
3	All Three Observers	P_1^3
3	Two or Three Observers	$P_1^3 + 3P_1^2Q_1$
3	One or More Observers	$P_1^3 + 3P_1^2Q + 3P_1Q_1^2$

TABLE 48

PERFORMANCE OF INDIVIDUAL OPERATORS AND OF VARIOUS TEAMS OF INDEPENDENT OBSERVERS

Number of Observers	Decision Rule, i.e., What Responses Shall be Recorded or Counted	Detection Per Cent	Proportion of all Responses that are to Nontargets
1	All Responses	23.8	.779
2	Only Objects* Found by Both	12.1	.667
2	Objects Found by Either or Both	35.6	.851
3	Only Objects Found by all Three	8.9	.688
3	Objects Found by Two or Three	20.5	.719
3	Objects Found by One or More	40.3	.824

* An object is any radar return (see "Explanation of Terms" section) that is identified by the observer as a target.

From this table the following conclusions may be drawn about using teams containing one or more independent observer with the radar pictures utilized in the present study:

1. **The percentage of targets present that are detected increases with the number of observers only when responses made by one or more team members are counted.**

2. **The detection percentage decreases with increase in team size when targets are counted only when found by all team members.**

3. **Under the one-or-more rule, the proportion of all responses that represent nontargets identifed as targets (the proportion of false positives) is larger with either 2 or 3 man teams than for the one-man team or operator.**

4. **The proportion of false positives is smaller with both two and three man teams when only unanimous responses are counted.**

5. **If there are many more targets than there is ammunition, then the two-member team in which only those objects designated by both observers as targets are attacked offers the highest probability that the attack will be upon a real target, i.e., that the percentage of false positives is a minimum.**

6. **Compared to one observer, teams of independent observers that utilize the "one or more" rule will detect an appreciably larger percentage of the targets, and will have only a small increase in the proportion of all responses that represent nontargets designated as targets.**

It is apparent that very little would be gained by the use of multiple operators working independently at the target location task on the radar film strip used in this study. Indeed, from an inspection of the binomial expansion, it is obvious that when target detection percentages are low and false positives are frequent, little gain can be expected from using multiple independent operators. However, this does not rule out the possibility that cooperating observers, in contrast to independent observers, may be able to improve system capability.

M. AGREEMENT COEFFICIENTS

It is obvious that radar observers differ in both the number of targets that they detect and in the number of nontarget objects that they mistake for targets. It is also to be expected that not all observers detect the same targets or respond to the same false positives, even when the actual numbers of the two classes of objects are the same for the two observers. It is also clear that some targets are easily detected, leading to relatively large numbers of observers detecting them: they are "popular" targets. Difficult targets, with low detection probabilities, would be "unpopular" targets. It is to be expected that some observers would have a higher portion of popular targets than would others. The same would hold true of false positives.

These speculations prompt one to ask the question: "How do the performance measures of observers relate to the "popularity" (or response frequencies or probability of detection) of the objects to which they respond?" To answer this question requires an index or measure which will permit comparison of their object choices to the choices of the whole group of observers. A simple and direct measure is the average relative response frequency or detection probability of all of the targets that the observer detects. A similar measure could be taken for the false positives responded to by the observer. For an individual observer the portion of all observers who detect each target to which the observer also responds may be added and the sum divided by the number of targets; in short, the sum of response probabilities of detected targets divided by the number of detected targets. The same procedure may be used for false positives. Note that such an index can not exceed unity for either class of object.

Table 49 lists, for each of the 20 observers, the coefficients of agreement for targets, for false positives, for the sum of the two coefficients, and some performance measures for comparisons with agreement coefficients. Due to the large numbers of false positives and the resultant huge amount of work involved in examining every one of the hundreds of data photographs taken of the display at every aircraft speed, the table was prepared for responses made only at a simulated aircraft speed of 700 knots.

Note that the two observers with the smallest number of detections, individuals "C" and "J", had relatively high agreement coefficients for detections, while the two observers with the greatest number of detections ("L" and "M") had relatively low coefficients. Also, the observer with the greatest number of false positives, "D", had the lowest coefficient for detections and also the lowest coefficient for false positives. The second to highest number of false positives was for an observer "E" with the next to lowest coefficient for false positives. Also, observer "R", with the next to lowest number of detections, had the highest coefficient for false positives. Observer "S", with the next to highest detection agreement coefficient, had the next to highest agreement coefficient for false positives. From these observations it appears likely that numbers of targets detected and numbers of false positives responded to are related, possibly quite highly, to the agreement coefficients. The magnitude of the relationships are given by the correlation coefficients listed in table 50.

Examination of the correlation table reveals that: (1) Number of targets detected correlates negatively with both agreement coefficients. Those who tend to find many targets have an above average proportion of responses that are made to targets with low detection probabilities and also a higher than average proportion of low response probability ("unpopular") false positives. Similarly, the responses of those who tend to find few targets tend to include in their responses an above average proportion of both popular targets and popular nontargets. (2) Number of false positives also correlates negatively with both agreement coefficients. Those who respond to many false positives tend to have an above average proportion of both unpopular targets and unpopular false positives. To do well on number of false positives, one must not respond to unpopular or

TABLE 49

AGREEMENT COEFFICIENTS FOR DETECTIONS AND FOR FALSE POSITIVES AND SOME PERFORMANCE MEASURES FOR COMPARISON

	Detections		False Positives		Percentage of False Positives		Agreement Coefficient, C					
							Detections		False Positives		C	Rank on
Observer	Number+	Rank	Number+	Rank	%F++	Rank	Number	Rank	Number	Rank	Sum	Sum
A	20	7	68	14	77.27	14	.5225	10	.2861	7	.8086	8
B	17	10.5	47	10	73.44	8	.5912	17	.3490	11	.9402	15
C	12*	18.5	101	16	89.38	20	.5666	14	.2703	5	.8369	10
D	27	3	155**	20	85.16	18	.4204*	2	.2110*	2	.6314	1
E	24	5	146**	19	85.88	19	.4312	3	.2055*	1	.6367	2
F	17	10.5	39	7.5	69.64	5	.5176	9	.3513	13.5	.8689	11
G	14	16.5	25*	1	64.10	1	.6250**	20	.3513	13.5	.9763	18
H	15	14.5	38	6	71.70	6	.5300	11	.3790	16	.9090	13
I	15	14.5	57	13	79.17	16	.4634	6	.3491	12	.8125	9
J	21	6	39	7.5	65.00	2	.5429	12	.3808	17	.9237	14
K	10*	20	30*	2	75.00	10.5	.5950	18	.3734	15	.9684	16
L	30**	1	104	17	77.61	15	.4500	5	.3087	9	.7587	5
M	29**	2	97	15	76.98	13	.4362	4	.2650	4	.7012	4
N	16	12.5	48	11	75.00	10.5	.5813	16	.3198	10	.9011	12
O	26	4	112	18	81.16	17	.4096*	1	.2348	3	.6444	3
P	19	8	53	12	73.61	9	.4895	7	.2896	8	.7791	7
Q	16	12.5	35	4.5	68.63	4	.5563	13	.3915	18	.9478	17
R	12*	18.5	31	3	72.09	7	.5792	15	.4065**	20	.9857	19
S	18	9	35	4.5	66.04	3	6056**	19	.3986**	19	1.0042	20
T	14	16.5	43	9	75.44	12	.5036	8	.2710	6	.7746	6
Mean	18.60		65.15		75.11		.5209		.3196			
Median	17.00		47.50		75.00		.5264		.3344			
S.D.	5.84		39.66		6.85		.06756		.06274			

+Prorating could not be used for computing agreement coefficients, hence the means and standard deviations for number of detections and number of false positives will not agree with values in tables that utilize prorating.

++Prorating not used, so values do not correspond exactly with those in table 13.

*Lowest numbers in the column.

**Highest numbers in the column.

NOTE: The tabled values are for a simulated aircraft speed of 700 knots.

TABLE 50

RELATIONSHIPS AMONG AGREEMENT COEFFICIENTS AND MEASURES OF OBSERVER RESPONSE AT 700 KNOTS

Agreement Coefficients		Correlated Measures: Number of Targets	Number of F.P.	F.P. Percentage	C_T
Target, C_T	r	−.7664	−.7757	−.5919	+.7348
	R	+.6648	−.7676	−.6477	+.7004
False Positives, C_F	r	−.5808	−.8680	−.7878	
	R	+.4773	−.8721	−.8226	
Sum or $C_T + C_F$	r	−.7272	−.8807	−.7369	
	R	+.6246	−.9022	−.7906	

r = **Product** moment correlation coefficient

R = **Rank** correlation coefficient

NOTE: All r and R values have 18 degrees of freedom. All r values are statistically significant at the .01 level, as are many R values, although some only reach the .05 level. In column 1 the difference in algebraic sign of r and R values is attributable to the method of ranking e.g., the observer with the most target detections is ranked 1 on detections, while the person with the most false positives (F.P.'s) is ranked 20 on F.P., etc.

doubtful objects. This, of course, leads to poor performance on number of targets detected. (3) Both of the agreement coefficients and their sum or average are negatively correlated with the percentage of false positives. Those observers who obtain, as compared to the average observer, a high percentage of false positives tend to respond to a higher proportion of unpopular targets and a higher proportion of unpopular false positives. As indicated earlier in the present report, to keep the percentage of false positives low, i.e. keep accuracy high, one must be careful to avoid the less obvious, hence "unpopular", objects on the display. Such behavior leads to detecting a low percentage of the available targets. (4) The agreement coefficients for targets and for false positives are positively correlated. Whatever the number of targets detected, those observers who are high in the proportion of low popularity targets are also high in the proportion of low popularity false positives. Similarly, those low on proportion of popular targets are low on proportion of popular false positives. (5) Adding the coefficients for the two types of objects, which would yield exactly the same correlations with other measures as their average, leads to lower correlations with performance measures, but the correlations are still high and statistically significant. Averaging the coefficients does not appear to be worthwhile.

The concept of an index of response similarity or coefficient of agreement introduced in the present paper has provided some interesting insights into the radar object selection behavior of radar observers. It was initially hoped that this way of looking into object choices of observers might yield some cues on how to solve, or at least reduce the magnitude of, the false positive problem. This does not appear to be the case. The conventional measures of number of detections, number of false positives, percentage of false positives and reaction or response time are not improved for observer selection purposes by the addition of agreement coefficients.

REVIEW OF RESULTS

1. INDIVIDUAL DIFFERENCES IN OBSERVERS

On every measure of performance examined (number of targets detected, number of false positives, target travel, reaction time to objects reported as targets, and response accuracy) the differences between individuals were large to very large. On every measure of performance the differences between individual observers were greater than the differences between group averages at different aircraft speeds.

2. EFFECT OF AIRCRAFT SPEED ON OBSERVER PERFORMANCE

Tripling aircraft speed produced several statistically significant changes in observer performance: the number of targets detected decreased by 16%, the average time to detect targets decreased by 56%, the average response time to false positives decreased by 60%, and the average aircraft travel between the display of an object and observer response to it increased by 33% for targets and 22% for false positives. A 20% decrease in the number of false positive responses was not statistically significant.

3. RELATIONSHIP BETWEEN NUMBER OF DETECTIONS AND AIRCRAFT SPEED

The number (or percentage) of targets detected, Y, was related to the simulated aircraft speed, V, by a linear equation of the form $Y = A - BV$. Thus, the number of targets detected decreases linearly with increase in aircraft speed.

4. RELATIONSHIP BETWEEN NUMBER OF FALSE POSITIVES AND SPEED

The number of false positive decreases linearly with increase in aircraft speed, the equation being of the form $X = C - DV$ where X is the number of false positives, V is aircraft speed and C and D are positive constants.

5. TARGET DIFFICULTY

Most of the targets in the present study were difficult to find from examination of the radar display. For no type of target, even at the slowest aircraft speed, did the percentage of targets that were detected exceed 30% of those that were judged, prior to testing observers, to be detectable on the display. For example, at the maximum aircraft speed only 9% of the airfields were detected.

6. AIRCRAFT TRAVEL PRIOR TO DETECTION

The ground distance, S, traveled by the simulated aircraft between the appearance of a target upon the display and its detection by observers was related to aircraft speed, V, by a linear equation of the form $S = A + BV$. The same form of function related position on the screen, S, when detected, to aircraft speed. A tripling of aircraft speed was accompanied by an increase of only 30% in screen travel before detection.

7. PERCENTAGE OF DETECTIONS AT VARIOUS SCREEN POSITIONS

The percentage of targets detected, P, at any distance, X, down the display screen follows an exponential equation of the form $P = e^{A + BX}$, where A and B are positive constants.

Since B, the constant multiplier of X in this equation, is linearly related to aircraft speed, V, it follows that the percentage of targets detected is described by the equation $P=e^{A + CX - DVX}$, where A, C, and D are positive constants. This equation may be rearranged to yield $P = e^{A + (C - DV)X}$.

8. AVERAGE DETECTION TIME AND AIRCRAFT SPEED

The average time taken to detect targets decreased as the logarithm of aircraft speed, i.e., fell off slowly as speed increased. The average detection time, $\bar{t}$, is related to aircraft speed, V, by the equation $\bar{t} = B - A \operatorname{Log}(V)$, where A and B are positive constants. The "fit" of this equation to the data was excellent.

9. LOCUS OF LOST TARGETS AT HIGHER SPEEDS

The most serious loss of targets, i.e., targets that were not detected, at higher aircraft speeds occurred within the first 8 nautical miles of aircraft travel following the appearance upon the display screen of the target images. This loss was not regained further down the display.

10. FALSE POSITIVE TARGET SIGNATURES
False positives are numerous because the environment contained many objects whose radar signatures are not distinguishable from those of real targets. In fact, a large portion of the false targets have signatures more like those of "good" real targets than does the average real target. This fact emerged from a study of target signatures and signatures of nontargets mistaken by subjects for targets. Training of subjects could not be blamed for the bulk of the false positives, and in the future it appears that, for unbriefed targets, the problem of excessive false positives will not be solved by training alone. Higher radar resolution, better dynamic range of the display, and/or use of other sensors to supplement the radar may be required for unbriefed targets.

11. USE OF MULTIPLE RADAR OBSERVERS
The performance of teams of two and of three independent observers was calculated by using the binomial theorem. The detection probabilities of every target and of every one of the nontargets mistaken for targets were determined at the 700 knot speed so that the theorem could be used. Performance was calculated under various decision rules as to what would be counted as a target detection. The percentage of targets detected increased as team size increased when responses made by one or more observers were counted. Decision rules that slightly reduced the percentage of false positives drastically reduced the percentage of available targets that were detected. The problem of excessive numbers of false positives was not solved by using teams of independently-working observers. What teams of cooperating, rather than independently-working, observers could do was not examined in the present study.

12. NUMBER OF DETECTIONS AND NUMBER OF FALSE POSITIVES
The statistically significant positive correlation (+.67) between the number of targets detected and the number of non-targets mistaken for targets indicates that observers who were relatively "good" on either measure tended to be "poor" on the other. However, there were exceptions.

13. NUMBER OF DETECTIONS AND PERCENTAGE OF FALSE POSITIVES
While observers who detected more targets also tended to mistake more non-targets for targets, they also tended to have a lower percentage of false positives.

14. REACTION TIME VERSUS DETECTIONS AND FALSE POSITIVES
The rapidity with which observers detected targets was not significantly related to either the percentage of the targets that they detected nor to the percentage of all of their responses that were made to nontarget objects.

15. CONFIDENCE IN RESPONSE CORRECTNESS AND PERFORMANCE
For either targets or false positives the average expressed confidence of observers that objects designated by them as targets were indeed targets was closer to medium than to high confidence. Expressed confidence did not vary with aircraft speed. At every aircraft speed and for the over-all-speeds average, the average level of confidence in correctness of response for targets only slightly exceeded that for false positives. Over 40% of all observers were either more confident of incorrect choices than of correct choices or else very nearly as confident. The confidence of individuals for the two types of objects were positively correlated: there was a tendency for those highly confident for targets to be highly confident for false positives, etc.

From the above one may say that for unbriefed large SLR targets of the types used in the present study and with a similar radar, expressed observer confidence has little if any value in discriminating between targets and false positives. Expressed confidence has little utility in solving the problem of excess numbers of false positives.

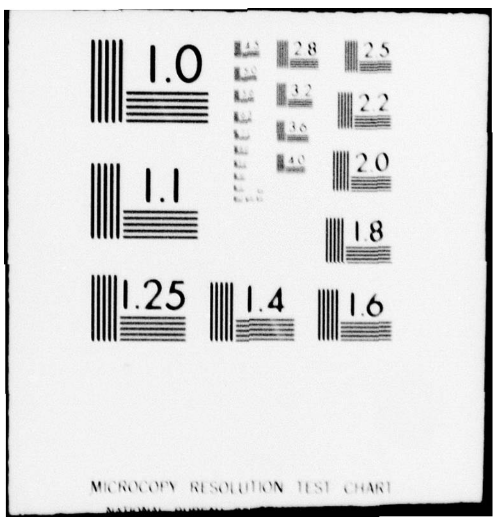
16. INDICE OF RESPONSE SIMILARITY
The average relative response frequency or detection probability of all of the targets detected by an observer is an indice of the similiarity of his target choices to those of other observers, and may be called an agreement coefficient. An analagous coefficient may be calculated for false positives. Both types of agreement coefficients were negatively correlated to a statistically significant degree with number of targets detected, number of false positives and percentage of false positives. Those who detect many targets designate many "unpopular"

AD-A060 908 AEROSPACE MEDICAL RESEARCH LAB WRIGHT-PATTERSON AFB OHIO F/G 5/9
THE BEHAVIOR OF OBSERVERS IN DETECTING UNBRIEFED TARGETS AT DIF--ETC(U)
JUN 78 H C SELF

UNCLASSIFIED AMRL-TR-77-95 NL

2 OF 2
AD
A060908

END
DATE
FILMED
1-79
DDC


1.0
2.8
2.5
3.2
2.2
3.6
4.0
2.0
1.1
1.8
1.25
1.4
1.6
MICROCOPY RESOLUTION TEST CHART

targets and nontargets; those who find many false positives tend to do likewise. Those who have a high percentage of false positives also tend to have a higher percentage of both unpopular targets and unpopular false positives. These findings are logical and are not surprising. The concept of response similarity or agreement coefficient does not appear to be a useful tool for supplementing the more common measures of performance in selecting superior observers, or in solving the false positive problem.

CONCLUSIONS AND RECOMMENDATIONS

1. TARGET DIFFICULTY

With high-resolution side-looking radar (SLR) similar in performance to the radar used in the present study, one should not expect to find and identify the majority of large unbriefed targets, such as airfields, dams, industry, railroad yards and tank farms. One can expect similar equipment to be of little value against small unbriefed targets, which would be even more difficult. These statements apply to real or near-real-time operations, rather than to situations in which considerable time is available for detailed examination of radar returns.

2. THE EFFECT OF AIRCRAFT SPEED ON DETECTION PERCENTAGE

There was a definite (statistically significant), but quite small, decrease of about 20% in the number or percentage of targets that were detected at 2110 knots as compared to 700 knots. With a radar similar in capability to the one used in the present study and with similar large unbriefed targets, it may be concluded that the number of targets detected at triple-sonic speeds should be almost as large as at sonic speeds.

3. THE FALSE POSITIVE PROBLEM

Users of high quality SLR have a severe problem in distinguishing between targets and nontarget objects, even in the case of large targets. Mistaking nontarget objects for targets is a severe problem because many objects on the terrain appear, on radar displays, to be targets, even to radar signature experts who take much time to study the displayed image. It follows that intensive training of observers will not, by itself, solve the false positive problem.

Research is required to determine if, and to what extent, increasing radar ground resolution and/or dynamic range of the recording and display can be useful in increasing the percentage of the targets that are detected and in reducing the number of objects that are mistaken for targets.

4. DETECTION AND RADAR SIGNATURES

From the foregoing discussion it follows that theoretical and empirical studies are required to determine why so many targets have radar signatures that are difficult or impossible to recognize as targets and why so many nontarget objects are confused with targets. One or more of the following factors are probably involved: material and construction of objects, "scintillation" effects due to the orientation of object surfaces relative to the aircraft, and the dynamic range and ground resolution of the radar.

5. SELECTION OF OBSERVERS

At any aircraft speed the range of performance of observers is typically 2:1 or more on most measures of performance. Data from repeated testing indicate that there is appreciable reliability in the ranking of observers: some are consistently superior on any one measure or even on more than one. Differences between observers on percentage of targets detected, accuracy of identification and speed or reaction are larger than differences attributable to a 3:1 change in aircraft speed. With large unbriefed targets and a high resolution SLR, the performance of the most efficient radar observers at triple-sonic aircraft speeds will appreciably exceed the performance of the least efficient at sonic speeds.

To considerably upgrade a near-real-time reconnaissance or recon/strike system, select only the most efficient observers.

6. TEAMS OF OBSERVERS

The response probabilities (relative response frequencies) of every target and of every object mistaken for a target were calculated from the test data at the 700 knot speed. These probabilities were used in the binomial theorem to predict the behavior of teams of two and three observers working independently under various decision rules on what would be designated as a target. It was shown that such teams of two or three observers would be little better than a one-man team, i.e., a lone observer, in solving the false positive problem. Decision

rules that slightly reduced the percentage of false positives drastically reduced the percentage of available targets that were detected.

Research is needed to determine if, and under what conditions, the false positive problem can be handled by the use of teams of cooperating, rather than independently-working, observers. It is not unlikely that such teams will also be of little value due to the nature of radar target signatures: signatures of nontarget and target objects are often too similar.

7. PREDICTION OF OBSERVERS BEHAVIOR

The large variability between observers and the appreciable variability within observers tends to conceal the regularity or lawfulness of search and detection behavior. Even so, sometimes rather simple mathematical equations describe or predict the target-finding behavior of the statistically average observer as a function of stimulus conditions. Some of these equations were formulated in the present study. Further research is needed to quantify the effects of variables other than those that were examined.

APPENDIX I

PRO-RATING OF DATA

At the start of every test session each subject was told to avoid obscuring the field of view of the data camera with his head or arm when he took a picture with it. Also, experimental subjects were cautioned at intervals during testing. However, an occasional picture was obscured so that performance data were not obtainable from an examination of the picture. Obscured pictures for each subject were divided into detections and false positives in proportion to the relative occurrence of these response types in the scorable pictures for that subject. These were then added to the scorable responses to obtain data more representative of the subject performance than would have been obtained if obscured pictures had not been included. This pro-rating procedure accounts for the fractional numbers of detections and false positives found in some of the tables in this report.

REFERENCES

Anonymous (Advertisement). *Aviation Week and Space Technology*, April 17, 1967, *86*, 32-33.

Baker, C. A., Morris, D. F. and Steedman, W. C., 1960. Target Recognition on Complex Displays. *Human Factors, Vol. 2, No. 2*, May 1960.

Bolin, S. F., Sadacca, R., and Martinek, H., 1965. *Team Procedures in Image Interpretation*. TR Note 164 (AD 480 533). U.S. Army Personnel Research Office. Support Systems Res. Lab., Washington, D.C.

Box, G. E. P. "Non-normality and Tests on Variance". *Biometrika*, 1953, *40*, 318-335.

Boynton, R. M. and Bush, W. R., 1955. *Laboratory Studies Pertaining to Visual Reconnaissance*. WADC Technical Report 55-304, Part I, (AD 91874), Wright Air Development Center, Wright-Patterson Air Force Base, Ohio.

Boynton, R. M., and Bush, W. R., 1957. *Laboratory Studies Pertaining to Visual Air Reconnaissance*. Wright Air Development Center Technical Report 55-304, Part II, (AD 118250), Wright-Patterson Air Force Base, Ohio.

Boynton, R. M., Elworth, C. and Palmer, R. M., 1958. *Laboratory Studies Pertaining to Visual Air Reconnaissance*. Wright Air Development Center Technical Report 55-304 (AD 142274), Wright-Patterson Air Force Base, Ohio.

Conklin, J. E., 1962. *Physical Parameters of Target-Background Complexity*. Hughes Aircraft Co., TIC 2732.20/113.

Edwards, A. L., 1950. Homogeneity of Variance and the Latin Square Design. *Psych. Bull. Vol. 47, No. 1*, Jan. 1950.

Edwards, Allen L., 1954, *Statistical Methods for the Behavioral Sciences*, Rinehart and Co., Inc., New York.

Edwards, A. E., 1960. *Experimental Design in Psychological Research*. Revised Edition, Rhinehart & Co., Inc., N.Y. pp. 254-264.

Elworth, C. L. 1964. *Target Detection as a Function of Display Speed*. Document No. D2-90544, The Boeing Co., Renton, Wash.

Grant, P. A., 1948. The Latin Square Principle in the Design and Analysis of Psychological Experiments. *Psych. Bull. Vol. 45, No. 5*, pp. 427-442, Sept. 1948.

Harger, R. O., 1969. *Synthetic Aperture Radar Systems: Theory and Design*. McGraw Hill, New York.

Hornseth, J. P., 1967. Individual and Two Man Team Target Finding Performance. *Human Factors*, Feb. 1967, *9*, 39-43. (Also published as AMRL-TR-66-127 (AD 624 487), Aerospace Medical Research Laboratory, Wright-Patterson Air Force Base, Ohio.

Jensen, H., Graham, L. C., Porcello, L. J., and Leith, E. N. Side-Looking Airborne Radar, *Scientific American*, Oct. 1977, Outside Front Cover, pp. 84-95.

McKechnie, D. F., 1967. *Effect of Briefing and Velocity on the Identification of Targets from Side-Looking Radar Imagery*, AMRL-TR-66-149 (AD 662 612), Aerospace Medical Research Laboratory, Wright-Patterson Air Force Base, Ohio.

Morrissette, J. O., Hornseth, J. P. and Shellar, K. Team Organization and Monitoring Performance. *Human Factors*, Vol. 17 (3), 296-300, 1975.

Nygaard, J. E., Slocum G. K., Thomas, J. O., Skeen, J. R., and Woodhull, J. G., 1964. *The Measurement of Stimulus Complexity in High-Resolution Sensor Imagery*. AMRL-TDR-64-29 (AD 603 007), Aerospace Medical Research Laboratory, Wright-Patterson Air Force Base, Ohio.

Olds, E. G., Mattson, T. B., and Odeh, R. E., 1956. *Notes on the Use of Transformations in the Analysis of Variance*. TR 56-308 (AD 97208). Wright Air Development Center, Wright-Patterson Air Force Base, Ohio.

Rhodes, F., 1964. *Predicting the Difficulty of Locating Targets from Judgments of Image Characteristics*. AMRL TDR 64-119 (AD 601 375), Aerospace Medical Research Laboratory, Wright-Patterson Air Force Base, Ohio.

Rhodes, F. and Self, H. C., 1964. *The Effect of Direction and Speed of Image Motion Upon Target Detection with Side-Looking Radar*. AMRL TDR 64-45 (AD 603 598), Aerospace Medical Research Laboratory, Wright-Patterson Air Force Base, Ohio.

Roetling, P. G., Hammil, H. B., and Holliday, T. M., 1963. *Quality Categorization of Aerial Reconnaissance Photography*. RADC TDR 63-297, Rome Air Development Center, Griffis Air Force Base, New York.

Schafer, T. H., Jan. 1947. *Detection of a Signal by Several Observers*. U.S. Naval Electronics Report No. 101 (AD 178821).

Self, H. C., McKechnie, D. F., Van Ausdall, B. A., and Welch, J. C., 1963 (Unclassified Title). *Side-Looking Radar Training Material for the 665A Program* [redacted] AMRL Memorandum P-64 (AD 516 496), Aerospace Medical Research Laboratory, Wright-Patterson Air Force Base, Ohio.

Self, H. C. and Rhodes, F., 1964. *The Effect of Simulated Aircraft Speed on Detecting and Identifying Targets from Side-Looking Radar Imagery*. AMRL TDR 64-60 (AD 603 014), Aerospace Medical Research Laboratory, Wright-Patterson Air Force Base, Ohio.

Self, H. C., 1972. "Performance Measures, Observer Selection, and Reconnaissance/Strike Effectiveness," *Office of Naval Research Target Acquisition Symposium*, Naval Training Devices Center, Florida, 14-16 November 1972.

Siegel, S., 1956. *Nonparametric Statistics for the Behavioral Sciences*. McGraw-Hill Book Co., Inc., New York.

Snedecor, G. W., 1956. *Statistical Methods*. The Iowa State U. Press, Ames, Iowa.

Steedman, W. C. and Baker, C. A., 1960. Target Size and Visual Recognition. *Human Factors, Vol. 2, No. 2*, August 1960.

Van Ausdall, B. A., and Self, H. C., 1964. *Effects of Display Polarity on Target Detection with Side-Looking Radar*. AMRL TR 64082 (AD 609 246), Aerospace Medical Resarch Laboratory, Wright-Patterson Air Force Base, Ohio.

Whiteside, T. C. D., Oct. 1957. *Target Detection and Number of Observers*. FPRC 1022 (AD 211 028), Flying Personnel Research Committee, Air Ministry (England).

Wiener, E. L., Performance of Multi-man Monitoring Teams. *Human Factors*. *Vol. 6*, 179-184, 1964.

Williams, A. C., Jr., Simon, C. W., Haugen, R., and Roscoe, S. N., 1960. *Operator Performance in Strike Reconnaissance*, WADD TR 60-521 (AD 246 545), Wright Air Development Division, Wright-Patterson Air Force Base, Ohio.

Williams, L. G., and Borow, M. S. The Effect of Rate and Direction of Display Movement Upon Visual Search. *Human Factors*, April 1963, *5(2)*, 139-146.

Winer, B. J., 1962. *Statistical Principles in Experimental Design*. McGraw Hill Book Co., New York.

Zeidner, J., Oct. 1962. *Some Human Factors Studies in the Army's Command Control Information Systems*. A paper presented at the Tenth Military Operations Research Symposium, Santa Monica, Calif.

NOTE: Those documents in the above listing which may be purchased from the National Technical Information Service are identified by a six digit AD number. See inside front cover of this report.

☆ U.S.Government Printing Office: 1978 — 757-080/46